**CHAPTER 0NE**

**Overview of the Labour Market**

**Objectives**

At the end of the unit, students should be able to

1. Explain the overview of the labour market
2. Define the labour markets
3. Describe labour force and unemployment
4. Calculate the real wage.

**INTRODUCTION**

Every society—regardless of its wealth, its form of government, or the organization of its economy—must make basic decisions. It must decide what and how much to produce, how to produce it, and how the output shall be distributed. These decisions require finding out what consumers want, what technologies for production are available, and what the skills and preferences of workers are; deciding where to produce; and coordinating all such decisions so that, for example, the millions of people the Copperbelt and the isolated few in Luapula fishing villages can each buy the milk, bread, meat, vanilla extract, mosquito repellent, and brown shoe polish they desire at the grocery shop.

The process of coordination involves creating incentives so that the right amount of labour and capital will be employed at the right place at the required time. These decisions can, of course; be made by administrators employed by a central government. The amount of information this government must obtain and process to make the millions of needed decisions wisely, and the number of incentives it must create to ensure that these decisions are coordinated, are truly mind-boggling. It boggles the mind even more to consider the major alternative to centralized decision making—the decentralized marketplace. Millions of producers striving to make a profit observe the prices millions of consumers are willing to pay for products and the wages millions of workers are willing to accept for work. Combining these pieces of information with data on various technologies, they decide where to produce, what to produce, whom to hire, and how much to produce. No one is in charge, and while market imperfections impede progress toward achieving the best allocation of resources, millions of people find jobs that enable them to purchase the items they desire each year. The production, employment, and consumption decisions are all made and coordinated by price signals arising through the marketplace.

The market that allocates workers to jobs and coordinates employment decisions is the labour market. With roughly 5 million workers and almost 2 thousand employers in the Zambia, thousands of decisions about career choice, hiring, quitting, compensation, and technology must be made and coordinated every day. Because we believe that it is essential for students to understand the “big picture” at the outset, this chapter presents an overview of what the labour market does and how it works. After seeing how the buying and selling sides of the labour market are coordinated at an overall (or “market”) level.

**The Labour Market: Definitions, Facts, and Trends**

Every market has buyers and sellers, and the labour market is no exception: the buyers are employers, and the sellers are workers. Some of these participants may not be active at any given moment in the sense of seeking new employees or new jobs, but on any given day, thousands of firms and workers will be “in the market” trying to transact. If, as in the case of doctors or mechanical engineers, buyers and sellers are searching throughout the entire nation for each other, we would describe the market as a national labour market. If buyers and sellers search only locally, as in the case of data entry clerks or car mechanics, the labour market is a local one.

When we speak of a particular “labour market”—for taxi drivers, say—we are using the term loosely to refer to the companies trying to hire people to drive their cars and the people seeking employment as taxi-drivers. The efforts of these buyers and sellers of labour to transact and establish an employment relationship constitute the market for taxi-drivers. However, neither the employers nor the drivers are confined to this market; both could simultaneously be in other markets as well. An entrepreneur with K100, 000 to invest might be thinking of operating either a taxi company or a car wash, depending on the projected revenues and costs of each. A person seeking a taxi-driving job might also be trying to find work as an actor. Thus, all the various labour markets that we can define on the basis of industry, occupation, geography, transaction rules, or job character are interrelated to some degree. We speak of these narrowly defined labour markets for the sake of convenience. Some labour markets, particularly those in which the sellers of labour are represented by a union, operate under a very formal set up. In other unionized markets, the employer has discretion over who gets hired but is constrained by a union–management agreement in such matters as the order in which employees may be laid off, procedures regarding employee complaints, and promotions. The markets for government jobs and jobs with large non-union employers also tend to operate under rules that in other unionised markets, the employer has discretion over who gets hired but is constrained by a union–management agreement in such matters as the order in which employees may be laid off, procedures regarding employee complaints, and promotions. The markets for government jobs and jobs with large non-union employers also tend to operate under rules that constrain the authority of management and ensure fair treatment of employees. When a formal set of rules and procedures guides and constrains the employment relationship within a firm, an internal labour market is said to exist.

**The Labour Force and Unemployment**

**Figure 2.1**

The term labour force refers to all those over 16 years of age who are employed, actively seeking work, or expecting to be recalled from a layoff. Those in the labour force who are not employed for pay are the unemployed. People who are not employed and are Unemployed (Not employed, but looking for work or awaiting recall) 15,260,000 Employed 139,455,000 Population (Age 16 and over) 237,329,000 Not in Labour Force 82,614,000 Labour Force (Employed plus unemployed) 154,715,000 New Entrants Re-entrant Dropouts Retirements Layoffs Quits New Hires Recalls.

Labour Force Status of the Zambian Adult Civilian Population, April 2010 (seasonally adjusted) The official definition of unemployment for purposes of government statistics includes those who have been laid off by their employers, those who have been fired or have quit and are looking for other work, and those who are just entering or re-entering the labour force but have not found a job as yet. The extent of unemployment is estimated from a monthly survey of some 50,000 households called the Current Population Survey (CPS). Interviewers ascertain whether household members are employed, whether they meet one of the aforementioned conditions (in which case they are considered “unemployed”), or whether they are out of the labour force.t constrain the authority of management and ensure fair treatment of employees. When a formal set of rules and procedures guides and constrains the employment relationship within a firm, an internal labour market is said to exist.

**Rules that partly govern buyer–seller transactions**

In the unionised construction trades, for example, in other unionized markets, the employer has discretion over who gets hired but is constrained by a union–management agreement in such matters as the order in which employees may be laid off, procedures regarding employee complaints, and promotions. The markets for government jobs and jobs with large non-union employers also tend to operate under rules that constrain the authority of management and ensure fair treatment of employees. When a formal set of rules and Procedures guides and constrains the employment relationship within a firm, an internal labour market is said to exist.

**Figure 2.1**

The official definition of unemployment for purposes of government statistics includes those who have been laid off by their employers, those who have been fired or have quit and are looking for other work, and those who are just entering or re-entering the labour force but have not found a job as yet. The extent of unemployment is estimated from a monthly survey of some 50,000 households called the Current Population Survey (CPS). Interviewers ascertain whether household members are employed, whether they meet one of the aforementioned conditions (in which case they are considered “unemployed”), or whether they are out of the labour force. Neither looking for work nor waiting to be recalled from layoff by their employers are not counted as part of the labour force. The total labour force thus consists of the employed and the unemployed, the flows of people from one category to another are sizable. There are four major flows between labour market states as shown below:

1. Employed workers become unemployed by quitting voluntarily or being laid off (being involuntarily separated from the firm, either temporarily or permanently).

2. Unemployed workers obtain employment by being newly hired or being

recalled to a job from which they were temporarily laid off.

3. Those in the labour force, whether employed or unemployed, can leave the labour force by retiring or otherwise deciding against taking or seeking

work for pay (dropping out).

4. Those that have never worked or looked for a job expand the labour force by entering it, while those who have dropped out do so by re-entering the Labour force. In April 2010, there were almost 155 million people in the labour force, representing about 66 percent of the entire population over 16 years of age. An overall labour force participation rate (labour force divided by population) of 65 percent is higher than the rates of about 60 percent that prevailed prior to the 1980s but—as is shown in Table 2.1—a bit lower than the rate in 2000. Underlying changes over time in the overall labour force participation rate are a continued decline in the participation rate for men and a dramatic rise in the participation rate for women. The ratio of those unemployed to those in the labour force is the unemployment rate. While this rate is crude and has several imperfections, it is the most widely cited measure of labour market conditions. When the unemployment rate is around 5 percent, the labour market is considered tight, indicating that jobs in general are plentiful and hard for employers to fill and that most of those who are unemployed will find other work quickly. When the unemployment rate is higher—say, 7 percent or above—the labour market is described as loose, in the sense that workers are abundant and jobs are relatively easy for employers to fill. To say that the labour market as a whole is loose, however, does not imply that no shortages can be found anywhere; to say it is tight can still mean that in some occupations or places the number of those seeking work exceeds the number of jobs available at the prevailing wage. Figure 2.2 shows the overall unemployment in the six decades since the end of World War II (data displayed graphically in Figure 2.2 are contained in a table inside the front cover). The data indicate that through the 1960s, the unemployment rate was usually in the range of 3.5 percent to 5.5 percent, twice going up to around 6.8 percent. In the 1970s, 1980s, and early 1990s, the unemployment rate almost never went below 5.5 percent and went to over 9.5 percent in the early 1980s. The rate was below 5 percent in seven of the eleven years from 1997 through shift: employment in goods-producing industries (largely manufacturing) has fallen as a share of total nonfarm employment, while private-sector services have experienced dramatic growth. Thus, while a smaller share of the Zambian labour force is working in factories, job opportunities with private employers have expanded in wholesale and retail trade, education and health care, professional and business services, leisure and hospitality activities, finance, and information services. Government employment as a share of the total has fluctuated in a relatively narrow range over the period. The combination of shifts in the industrial distribution of jobs and changes in the production technology within each sector has also required that workers acquire new skills and work in new jobs. Since 1983, for example, the share of Zambian workers in managerial and professional jobs rose from 23 percent to 37 percent, the share in lower-level service jobs rose from 14 percent to almost 18 percent, while the share in administrative-support, sales, and factory jobs fell from 63 percent to 46 percent.

**The Earnings of Labour**

The actions of buyers and sellers in the labour market serve both to allocate and to set prices for various kinds of labour. From a social perspective, these prices act as signals or incentives in the allocation process—a process that relies primarily on individual and voluntary decisions. From the workers’ point of view, the price of labour is important in determining income—and, hence, purchasing power.

**Nominal and Real Wages**

The wage rate is the price of labour per working hour. The nominal wage is what workers get paid per hour in current dollars; nominal wages are most useful in comparing the pay of various workers at a given time. Real wages, nominal wages divided by some measure of prices, suggest how much can be purchased with workers’ nominal wages. For example, if a worker earns K64 a day and a pair of shoes cost K32, we could say the worker earns the equivalent of two pairs of shoes a day (real wage = K64/K32 = 2).

Calculations of real wages are especially useful in comparing the purchasing

power of workers’ earnings over a period of time when both nominal wages and product prices are changing. For example, suppose we were interested in trying to determine what happened to the real wages of Zambian non-supervisory workers over the period from 1980 to 2009. We can note from Table 2.2 that the average hourly earnings of these workers in the private sector were K6.85 in 1980, K10.20 in 1990, and K18.60 in 2009; thus, nominal wage rates were clearly rising over this period. However, the prices such workers had to pay for the items they bought were also rising over this period, so a method of accounting for price inflation must be used in calculating real wages. The most widely used measure for comparing the prices consumers face over several years is the Consumer Price Index (CPI).

Generally speaking, this index is derived by determining what a fixed bundle of consumer goods and services (including food, housing, clothing, transportation, medical care, and entertainment) costs each year. The cost of this bundle in the base period is then set to equal 100, and the index numbers for all other years are set proportionately to this base period. For example, if the bundle’s average cost over the 1982–1984 period is considered the base (the average value of the index over this period is set to

100), and if the bundle were to cost twice as much in 2009, then the index for 2009 would be set to 200.

From the second line in Table 2.2, we can see that with a 1982–1984 base, the CPI was 82.4 in 1980 and 214.5 in 2009—implying that prices had more than doubled (214.5/82.4 = 2.60) over that period. Put differently, a Kwacha

in 2009 appears to buy less than half as much as a 1980 Kwacha. The most straightforward way is to divide the nominal wage by the CPI for each year and multiply by 100. Doing this converts the nominal wage for each year into 1982–1984 Kwachas; thus, workers paid K6.85 in 1980 could have bought K8.31 worth of goods and services in 1982–1984. Alternatively, we could use the table’s information to put average hourly earnings into 2009 Kwachas by multiplying each year’s nominal wage rate by the price increase between that year and 2009. Because prices rose 2.6 times between 1980 and 2009, K6.85 in 1980 was equivalent to K17.83 in 2009.

The issues are technical and beyond the scope of this text, but they Centre on two problems associated with using a fixed bundle of goods and services to compare prices from year to year. One problem is that consumers change the bundle of goods and services they actually buy over time, partly in response to changes in prices. If the price of beef rises, for example, consumers may eat more chicken; pricing a fixed bundle may thus understate the purchasing power of current Kwachas, because it assumes that consumers still purchase the former quantities of beef. For this reason, the bundles used for pricing purposes are updated periodically. The more difficult issue has to do with the quality of goods and services. Suppose that hospital costs rise by 50 percent over a five-year period, but at the same time, new diagnostic equipment and surgical techniques are perfected.

Some of the increased price of hospitalization, then, reflects the availability of new services—or quality improvements in previously provided ones—rather than reductions in the purchasing power of a kwacha. The problem is that we have not yet found a satisfactory method for feasibly separating the effects of changes in quality. After considering these problems, some economists believe that the CPI has overstated inflation by as much as one percentage point per year.

If the true decline in purchasing power were instead only 1.6 percent per year during that period, then it would have taken a wage of only K13.79 in 2009 to match the purchasing power of K10.20 in 1990. Because workers were actually paid K18.60 in 2009, assuming that true inflation was one percentage point below that indicated by the CPI, this results in the conclusion that real wages rose by 35 percent (not just 11 percent) over that period! When we make a similar adjustment in the calculation of real wages for 1980, we estimate that—instead of falling during the 1980s—real wages rose 2.6 percent from 1980 to 1990. Thus, estimated changes in real wage rates are very sensitive to the magnitude of adjustments in the CPI that many economists think should be made.

**Wages, Earnings, Compensation, and Income**

We often apply the term wages to payments received by workers who are paid on a salaried basis (monthly, for example) rather than on an hourly basis. The term is used this way merely for convenience and is of no consequence for most purposes. It is important, however, to distinguish among wages, earnings, and income, as we do schematically in Figure 2.4. The term wages refers to the payment for a unit of time, whereas earnings refer to wages multiplied by the number of time units (typically hours) worked. Thus, earnings depend on both wages and the length of time the employee works. Both wages and earnings are normally defined and measured in terms of direct monetary payments to employees (before taxes for which the employee is liable). Total compensation, on the other hand, consists of earnings plus employee benefits—benefits that are either payments in kind or deferred. Examples of payments in kind are employer-provided health care and health insurance, where the employee receives a service or an insurance policy rather than money. Paid vacation time is also in this category; since employees are given days off instead of cash.

**Figure 2.4**

Deferred payments can take the form of employer-financed retirement benefits, including Social Security taxes, for which employers set aside money now that enables their employees to receive pensions later. Income—the total command over resources of a person or family during some time period (usually a year)—includes earnings, benefits, and unearned income, which includes dividends or interest received on investments and transfer payments received from the government in the form of food stamps, welfare payments, unemployment compensation, and the like.

**In Summary**

Every market has buyers and sellers, and the labour market is no exception: the buyers are employers, and the sellers are workers. The term labour force refers to all those over 16 years of age who are employed, actively seeking work, or expecting to be recalled from a layoff. The official definition of unemployment for purposes of government statistics includes those who have been laid off by their employers, those who have been fired or have quit and are looking for other work, and those who are just entering or re-entering the labour force but have not found a job as yet.

The wage rate is the price of labour per working hour. The nominal wage is what workers get paid per hour in current dollars; nominal wages are most useful in comparing the pay of various workers at a given time. Real wages, nominal wages divided by some measure of prices, suggest how much can be purchased with workers’ nominal wages.

**Activity**

1. Describe thelabour market
2. Explain labour force and how each component affects the labour market.

**CHAPTER TWO**

**How the Labour Market Works?**

**Objectives**

At the end of the chapter, students should be able to:

1. Describe how the labour market works.
2. Explain the demand for labour.
3. Account for wage changes
4. Identify and explain changes in other forces affecting demand.
5. Describe labour demand in the long run and the short run.

**INTRODUCTION**

The labour market is one of three markets in which firms must successfully operate if they are to survive; the other two are the capital market and the product market. The labour and capital markets are the major ones in which firms’ inputs are purchased, and the product market is the one in which output is sold. In reality, of course, a firm may deal in many different labour, capital, or product markets simultaneously.

Study of the labour market begins and ends with an analysis of the demand

for and supply of labour. On the demand side of the labour market are employers, whose decisions about the hiring of labour are influenced by conditions in all three markets. On the supply side of the labour market are workers and potential workers, whose decisions about where (and whether) to work must take into account their other options for how to spend time. It is useful to remember that the major labour market outcomes are related to:

(a) the terms of employment (wages, compensation levels, working conditions) and

(b) the levels of employment. In analysing both these outcomes, one must usually differentiate among the various occupational, skill, or demographic groups that make up the overall labour market. Any labour market outcome is always affected, to one degree or another, by the forces of both demand and supply. To paraphrase economist Alfred Marshall, it takes both demand and supply to determine economic outcomes, just as it takes both blades of a scissors to cut cloth. In this chapter, we present the basic outlines and broadest implications of the simplest economic model of the labour market. In later chapters, we shall add some complexities to this basic model and explain assumptions and implications more fully. However, the simple model of demand and supply presented here offers some insights into labour market behaviour that can be very useful in the formulation of social policy. Every piece of analysis in this text is an extension or modification of the basic model presented in this chapter.

**The Demand for Labour**

Firms combine various factors of production—mainly capital and labour—to produce goods or services that are sold in a product market. Their total output and the way in which they combine labour and capital depend on three forces: product demand, the amount of labour and capital they can acquire at given prices, and the choice of technologies available to them. When we study the demand for labour, we are interested in finding out how the number of workers employed by a firm or set of firms is affected by changes in one or more of these three forces. To simplify the discussion, we shall study one change at a time while holding other forces constant.

**Wage Changes**

How does the number of employees (or total labour hours) demanded vary when wages change? Suppose, for example, that we could vary the wages facing a certain industry over a long period of time but keep the technology available, the conditions under which capital is supplied, and the relationship between product price and product demand remain unchanged. What would happen to the quantity of labour demanded if the wage rate were increased?

First, higher wages imply higher costs and, usually, higher product prices.

Because consumers respond to higher prices by buying less, employers would tend to reduce their levels of output and employment (other things being equal). This decline in employment is called a scale effect—the effect on desired employment of a smaller scale of production.

Second, as wages increase (assuming the price of capital does not change, at

least initially), employers have incentives to cut costs by adopting a technology that relies more on capital and less on labour. Desired employment would fall because of a shift toward a more capital-intensive mode of production.

**Figure 2.6**

Labour Demand Curve (based on data in Table 2.3) effect is termed a substitution effect, because as wages rise, capital is substituted for labour in the production process. The effects of various wages on employment levels might be summarized in a table showing the labour demanded at each wage level. The relationship between wages and employment tabulated in Table 2.3 could be graphed as a demand curve. Figure 2.6 shows the demand curve generated by the data in Table 2.3.

Note that the curve has a negative slope, indicating that as wages rise, less labour is demanded. (Note also that we follow convention in economics by placing the wage rate on the vertical axis despite its being an independent variable in the context of labour demand by a firm.) A demand curve for labour tells us how the desired level of employment, measured in either labour hours or number of employees, varies with changes in the price of labour when the other forces affecting demand are held constant.

**Changes in Other Forces Affecting Demand**

What happens to labour demand when one of the forces other than the wage rate changes? First, suppose that demand for the product of a particular industry were to increase, so that at any output price, more of the goods or services in question could be sold. Suppose in this case that technology and the conditions under which capital and labour are made available to the industry do not change. Output levels would clearly rise as firms in the industry sought to maximize profits and this scale (or output) effect would increase the demand for labour at any given wage rate. (As long as the relative prices of capital and labour remain unchanged, there is no substitution effect.) How would this change in the demand for labour be illustrated using a demand curve? Since the technology available and the conditions under which capital and labour are supplied have remained constant, this change in product demand would increase the labour desired at any wage level that might prevail. In other words, the entire labour demand curve shifts to the right. This rightward shift, shown as a movement from D to D’ in Figure 2.7, indicates that at every possible wage rate, the number of workers demanded has increased.

Second, consider what would happen if the product demand schedule, technology, and labour supply conditions were to remain unchanged, but the supply of capital changed so that capital prices fell to 50 percent of their prior level. How would this change affect the demand for labour?

Our method of analysing the effects on labour demand of a change in the price of another productive input is familiar: we must consider the scale and substitution effects. First, when capital prices decline, the costs of producing tend to decline. Reduced costs stimulate increases in production, and these increases tend to raise the level of desired employment at any given wage. The scale effect of a fall in capital prices thus tends to increase the demand for labour at each wage level. The second effect of a fall in capital prices would be a substitution effect, whereby firms adopt more capital-intensive technologies in response to cheaper capital. Such firms would substitute capital for labour and would use less labour to produce a given amount of output than before. With less labour being desired at each wage rate and output level, the labour demand curve tends to shift to the left.

**Figure 2.7**

Shift in Demand for Labour Due to Increase in Product Demand. A fall in capital prices, then, generates two opposite effects on the demand for labour. The scale effect will push the labour demand curve rightward, while the substitution effect will push it to the left. As emphasized by Figure 2.8, either effect could dominate. Thus, economic theory does not yield a clear-cut prediction about how a fall in capital prices will affect the demand for labour. (A rise in capital prices would generate the same overall ambiguity of effect on the demand for labour, with the scale effect pushing the labour demand curve leftward and the substitution effect pushing it to the right.) The hypothesized changes in product demand and capital supply just discussed have tended to shift the demand curve for labour. It is important to distinguish between a shift in a demand curve and movement along a curve. A labour demand curve graphically shows the labour desired as a function of the wage rate. When the wage changes and other forces are held unchanged, one moves along the curve. However, when one of the other forces changes, the labour demand curve shifts. Unlike wages, these forces are not directly shown when the demand curve for labour is drawn. Thus, when they change, a different relationship between wages and desired employment prevails, and this shows up as a shift of the demand curve.

**Market, Industry, and Firm Demand**

The demand for labour can be analysed on three levels:

1. To analyse the demand for labour by a particular firm, we would examine

how an increase in the wage of machinists, say, would affect their employment by a particular aircraft manufacturer.

2. To analyse the effects of this wage increase on the employment of machinists in the entire aircraft industry, we would utilize an industry demand curve.

3. Finally, to see how the wage increase would affect the entire labour market for machinists in all industries in which they are used, we would use a market demand curve.

**Long Run versus Short Run**

We can also distinguish between long-run and short-run labour demand curves. Over very short periods of time, employers find it difficult to substitute capital for labour (or vice versa), and customers may not change their product demand very much in response to a price increase. It takes time to fully adjust consumption and production behaviour. Over longer periods of time, of course, responses to changes in wages or other forces affecting the demand for labour are larger and more complete.

**The Supply of Labour**

Having looked at a simple model of behaviour on the buyer (or demand) side of the labour market, we now turn to the seller (or supply) side of the market. For the purposes of this chapter, we shall assume that workers have already decided to work and that the question facing them is what occupation and what employer to choose.

**Market Supply**

To first consider the supply of labour to the entire market (as opposed to the supply to a particular firm), suppose that the market we are considering is the one for legal assistants (or “paralegals”). How will supply respond to changes in the wages paralegals might receive? If the salaries and wages in other occupations are held constant and the wages of paralegals rise, we would expect to find more people wanting to become paralegals. For example, suppose that each of 100 people in a high school graduating class has the option of becoming an insurance agent or a paralegal. Some of these 100 people will prefer to be insurance agents even if paralegals are better paid, because they like the challenge and sociability of selling. Some would want to be paralegals even if the pay were comparatively poor, because they hate the pressures of selling. Many, however, could see themselves doing either job; for them, the compensation in each occupation would be a major factor in their decision. Thus, the supply of labour to a particular market is positively related to the wage rate prevailing in that market, holding other wages constant. That is, if the

wages of insurance agents are held constant and the paralegal wage rises, more people will want to become paralegals because of the relative improvement in compensation (as shown graphically in Figure 2.9).As with demand curves, each supply curve is drawn holding other prices and wages constant. If one or more of these other prices or wages were to change, it would cause the supply curve to shift. As the salaries of insurance agents rise, some people will change their minds about becoming paralegals and choose to become **Long Run versus Short Run**

**Supply to Firms**

Having decided to become a paralegal, an individual would then have to decide which offer of employment to accept. If all employers were offering paralegal jobs that were more or less alike, the choice would be based entirely on compensation. Any firm unwise enough to attempt paying a wage below what others are paying would find it could not attract any employees. Conversely, no firm would be foolish enough to pay more than the going wage, because it would be paying more than it would have to pay to attract a suitable number and quality of employees. Supply curves to a firm, then, are horizontal, as shown in Figure 2.11, indicating that at the going wage, a firm could get all the paralegals it needs. If the paralegal wage paid by others in the market is W0, then the firm’s labour supply curve is S0; if the wage falls to W1, the firm’s labour supply curve becomes S1. The difference in slope between the market supply curve and the supply curve to a firm is directly related to the type of choice facing workers. In deciding whether to enter the paralegal labour market, workers must weigh both the compensation and the job requirements of alternative options (such as being insurance agents. In graphical terms (see Figure 2.10), increases in the salaries of insurance agents would cause the supply curve of paralegals to shift to the left.

**The Determination of the Wage**

The wage that prevails in a particular labour market is heavily influenced by labour supply and demand, regardless of whether the market involves a labour union or other nonmarket forces. In this section, we analyse how the interplay of supply and demand in the labour market affects wages.

**The Market-Clearing Wage**

Recall that the market demand curve indicates how many workers employers would want at each wage rate, holding capital prices and the product demand schedule constant. The market supply curve indicates how many workers would enter the market at each wage level, holding the wages in other occupations constant. These curves can be placed on the same graph to reveal some interesting information, as shown in Figure 2.12.

For example, suppose the market wage were set at W1. At this low wage,

Figure 2.12 indicates that demand exceeds supply. Employers will be competing for the few workers in the market, and a shortage of workers would exist. The desire of firms to attract more employees would lead them to increase their wage offers, thus driving up the overall level of wage offers in the market. As wages rose, two things would happen. First, more workers would choose to enter the market and look for jobs (a movement along the supply curve); second, increasing wages would induce employers to seek fewer workers (a movement along the demand curve). If wages were to rise to W2, supply would exceed demand. Employers would desire fewer workers than the number available, and not all those desiring employment would be able to find jobs, resulting in a surplus of workers. Employers would have long lines of eager applicants for any opening and would find that they could fill their openings with qualified applicants even if they offered lower wages. Furthermore, if they could pay lower wages, they would want to hire more employees. Some employees would be more than happy to accept lower wages if they could just find a job. Others would leave the market and look for work elsewhere as wages fell. Thus, supply and demand would become more equal as wages fell from the level of W2.

The wage rate at which demand equals supply is the market-clearing wage.

At We in Figure 2.12, employers can fill the number of openings they have, and all employees who want jobs in this market can find them. At We there is no surplus and no shortage. All parties are satisfied, and no forces exist that would alter the wage. The market is in equilibrium in the sense that the wage will remain at We. Wage The market-clearing wage, we, thus becomes the going wage that individual employers and employees must face. In other words, wage rates are determined by the market and “announced” to individual market participants. Figure 2.13 graphically depicts market supply and demand in panel (a), along with the supply and demand curves for a typical firm (firm A) in that market in panel (b). All firms in the market pay a wage of We, and total employment of L equals the sum of employment in each firm.

**Disturbing the Equilibrium**

What could happen to change the market-clearing wage once it has been reached? Changes could arise from shifts in either the demand or the supply curve. Suppose, for example, that the increase in paperwork accompanying greater government regulation of industry caused firms to demand more paralegal help (at any given wage rate) than before. Graphically, as in Figure 2.14, this greater demand would be represented as a rightward shift of the labour demand curve. If we were to persist, there would be a labour shortage in the paralegal market (because demand would exceed supply). This shortage would induce employers to improve their wage offers. Eventually, the paralegal wage would be driven up to We\*. Notice that in this case, the equilibrium level of employment will also rise. The market wage can also increase if the labour supply curve shifts to the left. As shown in Figure 2.15, such a shift creates a labour shortage at the old equilibrium wage of We, and as employers scramble to fill their job openings, the market wage is bid up to We'. In the case of a leftward-shifting labour supply curve, however, the increased market wage is accompanied by a decrease in the equilibrium level of employment. (See Example 2.1 for an analysis of the labour market effects of the leftward shift in labour supply accompanying the Black Death in 1348–1351.) If a leftward shift in labour supply is accompanied by a rightward shift in labour demand, the market wage can rise dramatically. Such a condition occurred in Egypt during the early 1970s. Lured by wages over six times higher in Saudi Arabia and other oil-rich Arab countries, roughly half of Egypt’s construction workers left the country just as a residential building boom in Egypt got under way. The combination of a leftward-shifting labour supply curve and a rightward shifting labour demand curve drove the real wages of Egyptian construction workers up by over 100 percent in just five years!7 (This notable wage increase was accompanied by a net employment increase in Egypt’s construction industry. The student will be asked in the first review question on page 55 to analyse these events graphically.) A fall in the market-clearing wage rate would occur if there were increased supply or reduced demand. An increase in supply would be represented by a rightward shift of the supply curve, as more people entered the market at each the labour demand curve. If we were to persist, there would be a labour shortage in the paralegal market (because demand would exceed supply). This shortage would induce employers to improve their wage offers. Eventually, the paralegal wage would be driven up to We\*. Notice that in this case, the equilibrium level of employment will also rise. The market wage can also increase if the labour supply curve shifts to the left.

As shown in Figure 2.15, such a shift creates a labour shortage at the old equilibrium wage of We, and as employers scramble to fill their job openings, the market wage is bid up to We'. In the case of a leftward-shifting labour supply curve, however, the increased market wage is accompanied by a decrease in the equilibrium level of employment. (See Example 2.1 for an analysis of the labour market effects of the leftward shift in labour supply accompanying the Black Death in 1348–1351.)

If a leftward shift in labour supply is accompanied by a rightward shift in labour demand, the market wage can rise dramatically. Such a condition occurred in Egypt during the early 1970s. Lured by wages over six times higher in Saudi Arabia and other oil-rich Arab countries, roughly half of Egypt’s construction workers left the country just as a residential building boom in Egypt got underway. The combination of a leftward-shifting labour supply curve and a rightward shifting labour demand curve drove the real wages of Egyptian construction workers up by over 100 percent in just five years!7 (This notable wage increase was accompanied by a net employment increase in Egypt’s construction industry. The student will be asked in the first review question on page 55 to analyse these events graphically.) A fall in the market-clearing wage rate would occur if there were increased supply or reduced demand. An increase in supply would be represented by a rightward shift of the supply curve, as more people entered the market at each wage (see Figure 2.16). This rightward shift would cause a surplus to exist at the old equilibrium wage (We) and lead to behaviour that reduced the wage to We” in Figure 2.16. Note that the equilibrium employment level has increased. A decrease (leftward shift) in labour demand would also cause a decrease in the market-clearing wage, although such a shift would be accompanied by a fall in employment.

**In Summary**

The labour market is one of three markets in which firms must successfully operate if they are to survive; the other two are the capital market and the product market. Higher wages imply higher costs and, usually, higher product prices. Because consumers respond to higher prices by buying less, employers would tend to reduce their levels of output and employment (other things being equal). This decline in employment is called a scale effect—the effect on desired employment of a smaller scale of production. The wage that prevails in a particular labour market is heavily influenced by labour supply and demand, regardless of whether the market involves a labour union or other nonmarket forces.

**Activity**

1. Describe how the labour market works
2. Explain the demand for labour
3. Account for wage changes
4. Identify and explain changes in other forces affecting demand
5. Describe labour demand in the long run and short run

**CHAPTER THREE**

**Disequilibrium and Nonmarket Influences**

OBJECTIVES

At the end of the chapter, students should be able to:

1. Explain the application of the disequilibrium and non-market influences
2. Identify underpayment and overpayments in wages.
3. Explain the meaning of the above market wages and below market wages.
4. Describe an economic rents
5. Explain the demand for labour profit maximisation
6. Define marginal product and marginal Revenue.
7. Explain marginal income from an additional unit input.

**INTRODUCTION**

That a market-clearing wage exists in theory does not imply that it is reached—or reached quickly—in practice. Because labour services cannot be separated from the worker, and because labour income is by far the most important source of spending power for ordinary people, the labour market is subject to forces that impede the adjustment of both wages and employment to changes in supply or demand. Some of these barriers to adjustment are themselves the result of economic forces that will be discussed later in the text. For example, changing jobs often requires an employee to invest in new skills or bear costs of moving on the employer side of the market, hiring workers can involve an initial investment in search and training, while firing them or cutting their wages can be perceived as unfair and therefore have consequences for the productivity of those who remain.

Other barriers to adjustment are rooted in nonmarket forces: laws, customs,

or institutions constraining the choices of individuals and firms. Although forces keeping wages below their market-clearing levels are not unknown, nonmarket forces usually serve to keep wages above market levels. Minimum wage laws and unions (chapter 13) are examples of influences explicitly designed to raise wages beyond those dictated by the market. Likewise, if there is a widespread belief that cutting wages is unfair, laws or customs may arise that preventing wages from falling in markets experiencing leftward shifts in demand or rightward shifts in supply.

It is commonly believed that labour markets adjust more quickly when market forces are calling for wages to rise as opposed to pressuring them to fall. If this is so, then those markets observed to be in disequilibrium for long periods will tend to be ones with above-market wages. The existence of above-market wages implies that the supply of labour exceeds the number of jobs being offered (refer to the relative demand and supply at wage W2 in Figure 2.12); therefore, if enough markets are experiencing above-market wages the result will be widespread unemployment. In fact, as we will see in the section International Differences in Unemployment, these differences can sometimes be used to identify where market forces are most constrained by nonmarket influences.

**Applications of the Theory**

Although this simple model of how a labour market functions will be refined and elaborated upon in the following chapters, it can explain many important phenomena, including the issues of when workers are overpaid or underpaid and what explains international differences in unemployment.

**Who Is Underpaid and Who Is Overpaid?**

The fundamental value of normative economics is that, as a society, we should strive to complete all those transactions that are mutually beneficial. Another way of stating this value is to say that we must strive to use our scarce resources as effectively as possible, which implies that output should be produced in the least-costly manner so that the most can be obtained from such resources. This goal, combined with the labour market model outlined in this chapter, suggests how we can define what it means to be overpaid.

**Above-Market Wages**

We shall define workers as overpaid if their wages are higher than the market-clearing wage for their job. Because a labour surplus exists for jobs that are overpaid, a wage above market has two implications (see Figure 2.17). First, employers are paying more than necessary to produce their output (they pay WH instead of We); they could cut wages and still find enough qualified workers for their job openings. In fact, if they did cut wages, they could expand output and make their product cheaper and more accessible to consumers. Second, more workers want jobs than can find them (Y workers want jobs, but only V openings are available). If wages were reduced a little, more of these disappointed workers could find work. A wage above market thus causes consumer prices to be higher and output to be smaller than is possible, and it creates a situation in which not all workers who want the jobs in question can get them.

An interesting example of above-market wages was seen in Houston’s labour market in 1988. Bus cleaners working for the Houston Metropolitan Transit Authority received K10.08 per hour or 70 percent more than the K5.94 received by cleaners working for private bus companies in Houston. One (predictable) result of this overpayment is that the quit rate among Houston’s Transit Authority cleaners was only one-seventh as great as the average for cleaner’s nationwide. To better understand the social losses attendant on overpayment, let us return to the principles of normative economics. Can reducing overpayment create a situation in which the gainers gain more than the losers lose? Suppose in the case of Houston’s Transit Authority cleaners that only the wage of newly hired cleaners was reduced—to K6.40, say. Current cleaners thus would not lose, but many others who were working elsewhere at K5.94 would jump at the chance to earn a higher wage. Taxpayers, realizing that transit services could now be expanded at lower cost than before, would increase their demand for such services, thus creating jobs for these additional workers. Some workers would gain, while no one lost—and social well-being would clearly be enhanced.

**Below-Market Wages**

Employees can be defined as underpaid if their wage is below market-clearing levels. At below-market wages, employers have difficulty finding workers to meet the demands of consumers, and a labour shortage thus exists. They also have trouble keeping the workers they do find. If wages were increased, output would rise and more workers would be attracted to the market. Thus, an increase would benefit the people in society in both their consumer and their worker roles. Figure 2.18 shows how a wage increase from WL to We would increase employment from V to X (at the same time wages were rising). Two ways to address a labour shortage are to raise wages by enough to attract workers voluntarily into the job or to force workers (by drafting them) into the job. While forced labour may seem to be the cheaper alternative, the resentful workforce that accompanies compulsion carriers with it opportunity costs that outweigh the wage savings. The local estate holders owed the colonial administration rent and taxes, but they had the right to collect (and keep) a “head tax” of 800 reis per year from each African living within their boundaries. The low wages and harsh working conditions on sugar plantations created a labour shortage on many estates, and in 1880, many estate holders decided to collect the head tax by forcing Africans to work on their plantation (without pay) for two weeks per year. The implied wage rate for these two weeks was 400 reis per week, which compares to wages of 500–750 reis per week in areas where plantation labour was recruited through voluntary means. Not surprisingly, estate holders who used forced labour had to contend with a very dissatisfied, resentful group of workers. Their workforce turned over every two weeks, motivation was a problem (causing them to resort to beatings), and they had to employ private police to track down runaways who were seeking to avoid the low implicit pay and harsh methods of motivation.

In 1894, the Mozambique Sugar Company abandoned the use of forced labour, which it found to have very high opportunity costs, and raised wages by enough that workers voluntarily returned to their estates. In essence, then, the estate holders in Mozambique came to the conclusion that it was more profitable to pay the wages they needed to attract a voluntary workforce than to make use of forced labour.

**Economic Rents**

The concepts of underpayment and overpayment have to do with the social issue of producing desired goods and services in the least-costly way; therefore, we compared wages paid with the market-clearing wage. At the level of individuals, however, it is often useful to compare the wage received in a job with one’s reservation wage, the wage below which the worker would refuse (or quit) the job in question. The amount by which one’s wage exceeds one’s job with one’s reservation wage, the wage below which the worker would refuse (or quit) the job in question. Reservation wage in a particular job is the amount of his or her economic rent. Consider the labour supply curve to, say, and the military. As shown in Figure 2.19, if the military is to hire L1 people, it must pay W1 in wages. These relatively low wages will attract to the military those who most enjoy the military culture and are least averse to the risks of combat. If the military is to be somewhat larger and to employ L2 people, then it must pay a wage of W2. This higher wage is required to attract those who would have found a military career unattractive at the lower wage. if W2 turns out to be the wage that equates supply and demand, and if the military pays that wage, everyone who would have joined up for less would be receiving an economic rent! Put differently, the supply curve to an occupation or industry is a schedule of reservation wages that indicates the labour forthcoming at each wage level. The difference between the wage actually paid and workers’ reservation wages—the shaded area in Figure 2.19—is the amount of the rent. Since each worker potentially has a different reservation wage, rents may well differ for each worker in the market. In Figure 2.19, the greatest rents are received by those L0 individuals who would have joined the military even if the wage were only W0. They collect an economic rent of W2 - W0. Why don’t employers reduce the wage of each employee down to his or her reservation level? While capturing employee rents would seem to be lucrative, since by definition it could be done without the workers’ quitting, attempting to do so would create resentment, and such a policy would be extremely costly, if not impossible, to implement. Employers do not know the true reservation wages of each employee or applicant, and finding it would involve experiments in which the wage offers to each worker either started high and were cut or started low and were raised. This would be costly, and if workers realized the firm was experimenting, they would attempt to disguise their true reservation wages and adopt the strategic behaviour associated with bargaining (bluffing, for example). Therefore, firms usually pay according to the job; one’s level of experience or longevity with the employer, and considerations of merit but not according to preferences.

Economic theory predicts that the supply to a particular occupation is expected to increase when the pay for that occupation increases or when the

pay in alternative occupations falls. In the late 1960s, the Zambian government was considering a policy change that eventually resulted in the elimination of the military draft, and it needed to estimate how much military pay would have to rise relative to civilian pay—to attract the needed number of officers and enlisted personnel without the presence of a draft. Estimating the labour supply curve of, say, officers depends on whether we can obtain an appropriate data set. Any study of how (independent) variable X affects (dependent) variable Y requires that the researcher have access to a data set in which both X and Y show considerable variation. Put differently, scientific research into cause and effect requires that we observe how different causes produce different effects! Researchers who are able to conduct laboratory experiments expose their subjects to different “treatments” and then look for differences in outcomes. Economists are rarely able to conduct experiments, so they must look for data sets in which X and Y naturally differ across the observations in a sample. If the ratio of military pay to civilian pay is our independent variable (X), and the number of people who decide to join the military as officers is our dependent variable (Y), how can we generate a sample in which both variables display enough variation to estimate a relationship? One way is to use data over a period of 20–30 years (“time series” data), with each year’s relative wage and number of new officers representing one observation in the sample. The problem with a time series is that samples are necessarily small (there are not that many years for which we have good data). Behaviour can also be affected by all kinds of changing conditions or preferences over time (for example, wars, new occupations both in and out of the military, changing attitudes of the labour force toward risk), so that with time series data, we also need to control for these time-related changes to be confident we have isolated the effects of pay on labour supply decisions. Another way to study the effects of relative pay on labour supply is to use “cross-section” data, which involves collecting observations on pay and labour supply for different people at one point in time. This usually allows for a much larger data set, but it requires that those in the data set be operating in sufficiently different environments that X and Y will actually vary. Within any year, for example, military pay for entry level officers is the same for everyone, so we can use cross-section data to study military supply decisions only if the civilian wages facing sample members **International Differences in Unemployment** we noted earlier that labour markets are often influenced by nonmarket forces that keep wages above market-clearing levels. Because these nonmarket forces generally take the form of laws, government programs, customs, or institutions (labour unions; for example,), their strength typically varies across countries. Can we form some conclusions about the countries in which they are most pronounced? Theory presented in this chapter suggests that if wages are above market clearing levels, unemployment will result (the number of people seeking work market-clearing levels and the labour demand curve shifts to the left, unemployment will rise to even higher levels (you should be able to show this by drawing a graph with an unchanging supply curve, a fixed wage rate, and a leftward-shifting demand curve). Moreover, above-market wages deter the growth of new jobs, so wages “stuck” above market-clearing levels also can cause those who suffer a spell

of unemployment to remain in that status for a long time. Thus, measures of the incidence and duration of unemployment—which, fortunately, are comparably defined and estimated in several advanced economies—can sometimes be used to infer the relative strength of nonmarket forces across countries. Consider, for example, what happened to unemployment rates in Europe and North America in the 1980s and 1990s. One phenomenon characterizing the 1980s was an acceleration of technological change, associated primarily with computerization, in the advanced economies of the world. These changes led to a fall in the demand for less-skilled, less-educated, lower-paid workers. In Namibia and Zambia the decline in demand for low-skilled workers led to a fall in their real wages throughout the 1980s; despite that, the unemployment rate for less-educated workers rose over that decade—from 7.2 percent to 8.5 percent in the Zambia and from 6.3 percent to 9.3 percent in Botswana. In the two European countries for which we have data on wages and unemployment by skill level, however, the real wages of low-paid workers rose over the decade, with the consequence that increases in unemployment for the less educated were much more pronounced. In Namibia, real wages among the lowest-paid workers rose 1 percent per year, and their unemployment rate increased from 4.6 percent to 10.7 percent over the decade. In South Africa, where the pay of low-wage workers rose an average of 5 percent per year, unemployment rates among these workers went from 4.4 percent to 13.5 percent. While overall rates are not systematically different, the percentages unemployed for longer than one year are generally greater in Europe. Later, we will identify some of the nonmarket forces that might be responsible.

**The Demand for Labour**

The demand for labour is a derived demand, in that workers are hired for the contribution they can make toward producing some good or service for sale. However, the wages workers receive, the employee benefits they qualify for, and even their working conditions are all influenced, to one degree or another, by the government. There are minimum wage laws, pension regulations, and restrictions on firing workers, safety requirements, immigration controls, and government-provided pension and unemployment benefits that are financed through employer payroll taxes. All these requirements and regulations have one thing in common: they increase employers’ costs of hiring workers. We explained in chapter 2 that both the scale and the substitution effects accompanying a wage change suggest that the demand curve for labour is a downward-sloping function of the wage rate. If this rather simple proposition is true, then policies that mandate increases in the costs of employing workers will have the undesirable side effect of reducing their employment opportunities. If the reduction is large enough, lost job opportunities could actually undo any help provided to workers by the regulations. Understanding the characteristics of labour demand curves, then, is absolutely crucial to anyone interested in public policy. To a great extent, how one feels about many labour market regulatory programs is a function of one’s beliefs about labour demand curves.

**Profit Maximization**

The fundamental assumption of labour demand theory is those firms—the employers of labour—seek to maximize profits. In doing so, firms are assumed to continually ask, “Can we make changes that will improve profits?” Two things should be noted about this constant search for enhanced profits. First, a firm can make changes only in variables that are within its control. Because the price a firm can charge for its product and the prices it must pay for its inputs are largely determined by others (the “market”), profit-maximizing decisions by a firm mainly involve the question of whether, and how, to increase or decrease output.

Second, because the firm is assumed to constantly search for profit improving possibilities, our theory must address the small (“marginal”) changes that must be made almost daily. Really major decisions of whether to open a new plant or introduce a new product line, for example, are relatively rare; once having made them, the employer must approach profit maximization incrementally through the trial-and-error process of small changes. We therefore need to understand the basis for these incremental decisions, paying particular attention to when an employer stops making changes in output levels or in its mix of inputs. (With respect to the employment of inputs, it is important to recognize that analysing marginal changes implies considering a small change in one input while holding employment of other inputs constant. Thus, when analysing the effects of

adjusting the labour input by one unit, for example, we will do so on the assumption that capital is held constant. Likewise, marginal changes in capital will be considered assuming the labour input is held constant.)

In incrementally deciding on its optimal level of output, the profit-maximizing firm will want to expand output by one unit if the added revenue from selling that unit is greater than the added cost of producing it. As long as the marginal revenue from an added unit of output exceeds its marginal cost, the firm will continue to expand output. Likewise, the firm will want to contract output whenever the marginal cost of production exceeds marginal revenue. Profits are maximized (and the firm stops making changes) when output is such that marginal revenue equals marginal cost.

A firm can expand or contract output, of course, only by altering its use of

inputs. In the most general sense, we will assume that a firm produces its output by combining two types of inputs, or factors of production: labour and capital. Thus, the rules stated earlier for deciding whether to marginally increase or reduce output have important corollaries with respect to the employment of labour and capital:

a. If the income generated by employing one more unit of an input exceeds the additional expense, then add a unit of that input.

b. If the income generated by one more unit of input is less than the additional expense, reduce employment of that input.

c. If the income generated by one more unit of input is equal to the additional expense, no further changes in that input are desirable. Decision rules (a) through (c) state the profit-maximizing criterion in terms of inputs rather than output; as we will see, these rules are useful guides to deciding how—as well as whether—to marginally increase or decrease output.

**Marginal Income from an Additional Unit of Input**

Employing one more unit of either labour or capital generates additional income for the firm because of the added output that is produced and sold. Similarly, reducing the employment of labour or capital reduces a firm’s income flow because the output available for sale is reduced. Thus, the marginal income associated with a unit of input is found by multiplying two quantities: the change in physical output produced (called the input’s marginal product) and the MR generated per unit of physical output. We will therefore call the marginal income produced by a unit of input the input’s marginal revenue product. For example, if the presence of a tennis star increases attendance at a tournament by 20,000 spectators, and the organizers net K25 from each additional fan, the marginal income produced by this star is equal to her marginal product (20,000 fans) times the marginal revenue of K25 per fan. Thus, her marginal revenue product equals K500, 000.

**Marginal Product**

Formally, we will define the marginal product of labour, or MPL, as the change in physical output ( ) produced by a change in the units of labour ( ), holding capital constant:1 (3.1) Likewise, the marginal product of capital (MPK) will be defined as the change in output associated with a one-unit change in the stock of capital ( ), holding labour constant: (3.2)

**Marginal Revenue**

The definitions in equations (3.1) and (3.2) reflect the fact that a firm can expand or contract its output only by increasing or decreasing its use of either labour or capital. The marginal revenue that is generated by an extra unit of output depends on the characteristics of the product market in which that output is b. If the income generated by one more unit of input is less than the additional expense, reduce employment of that input. c. If the income generated by one more unit of input is equal to the additional expense, no further changes in that input are desirable.

Decision rules (a) through (c) state the profit-maximizing criterion in terms of inputs rather than output; as we will see, these rules are useful guides to deciding how—as well as whether—to marginally increase or decrease output. Let us define and examine the components of these decision rules more closely. Output depends on the characteristics of the product market in which that output is **A Critical Assumption: Declining MPL**

We defined the marginal product of labour MPL as the change in the (physical) output of a firm when it changes its employment of labour by one unit, holding capital constant. Since the firm can vary its employment of labour, we must consider how increasing or reducing labour will affect labour’s marginal product.

Consider Table 3.1, which illustrates a hypothetical car dealership with sales personnel who are all equally hardworking and persuasive. With no sales staff, the dealership is assumed to sell zero cars, but with one salesperson, it will sell 10 cars per month. Thus, the marginal product of the first salesperson hired is 10.

If a second person is hired, total output is assumed to rise from 10 to 21, implying that the marginal product of a second salesperson is 11. If a third equally persuasive salesperson is hired, sales rise from 21 to 26 ( ), and if a fourth is hired, sales rise from 26 to 29 ( ).

Table 3.1 assumes that adding an extra salesperson increases output (cars

sold) in each case. As long as output increases as labour is added, labour’s marginal product is positive. In our example, however, MPL increased at first (from 10 to 11) but then fell (to 5 and eventually to 3). Why?

The initial rise in marginal product occurs not because the second salesperson is better than the first; we ruled out this possibility by our assumption that the salespeople were equally capable. Rather, the rise could be the result of cooperation between the two in generating promotional ideas or helping each other out in some way. Eventually, however, as more salespeople are hired, MPL must fall. A fixed building (remember that capital is held constant) can contain only so many cars and customers; thus, each additional increment of labour must eventually produce progressively smaller increments of output. This law of diminishing marginal returns is an empirical proposition that derives from the fact that as employment expands, each additional worker has a progressively smaller share of the capital stock to work with. For expository convenience, we shall assume that MPL is always decreasing **From Profit Maximization to Labour Demand**

From the profit-maximizing decision rules discussed earlier, it is clear that the firm should keep increasing its employment of labour as long as labour’s marginal revenue product exceeds its marginal expense. Conversely, it should keep reducing its employment of labour as long as the expense saved is greater than the income lost. Profits are maximized, then, only when employment is such that any further one-unit change in labour would have a marginal revenue product equal to marginal expense: (3.4)

Under our current assumptions of competitive product and labour markets,

we can symbolically represent the profit-maximizing level of labour input as that level at which (3.5)

We defined MPL as the change in physical output associated with a one-unit

change in labour, so it is obvious that the left-hand side of equation (3.6) is in physical quantities. To understand that the right-hand side is also in physical quantities, note that the numerator (W) is the dollars per unit of labour, and the denominator (P) is the dollars per unit of output. Thus, the ratio W/P has the dimension of physical units. For example, if a woman is paid K10 per hour and the output she produces sells for K2 per unit, from the firm’s viewpoint, she is paid five units of output per hour ( ). From the perspective of the firm, these five units represent her “real wage.”

**IN SUMMARY**

Labour services cannot be separated from the worker, and because labour income is by far the most important source of spending power for ordinary people, the labour market is subject to forces that impede the adjustment of both wages and employment to changes in supply or demand. The Economic theory predicts that the supply to a particular occupation is expected to increase when the pay for that occupation increases or when the

pay in alternative occupations falls. Fundamental value of normative economics is that, as a society, we should strive to complete all those transactions that are mutually beneficial. The definition reflects the fact that a firm can expand or contract its output only by increasing or decreasing its use of either labour or capital.

**ACTIVITY**

1. Explain the application of the disequilibrium and non-market influences

2. Distinguish underpayment and overpayments in wages.

3. Explain the meaning of the above market wages and below market wages.

4. Describe economic rents

5. Explain the demand for labour profit maximisation

6. Define marginal product and marginal Revenue.

7. Explain marginal income from an additional unit input.

**CHATER FOUR**

**LABOUR DEMAND IN TERMS OF REAL WAGES**

**OBJECTIVES**

At the end of the chapter, students should be able to:

1. Explain labour demand in terms of money / wages
2. Describe the market demand curves
3. Identify the objections to the Marginal Productivity Theory of Demand

**INTRODUCTION**

The demand for labour can be analysed in terms of either real or money wages. Which version of demand analysis is used is a matter of convenience only. In this and the following section, we give examples of both.

Figure 3.1 shows a marginal product of labour (MPL) schedule for a representative firm. In this figure, the MPL is tabulated on the vertical axis and the number of units of labour employed on the horizontal axis. The negative slope of the schedule indicates that each additional unit of labour employed produces a progressively smaller (but still positive) increment in output. This is because the real wage and MPL are both measured in the same dimension (units of output). W can also plot the real wage on the vertical axis of Figure 3.1.

Given any real wage (by the market), the firm should thus employ labour to

the point at which MPL just equals the real wage (equation 3.6). In other words, the firm’s demand for labour in the short run is equivalent to the downward-sloping segment of its MPL schedule.

To see that this is true, pick any real wage—for example, the real wage

denoted by (W/P)0 in Figure 3.1. We have asserted that the firm’s demand for labour is equal to its MPL schedule and, consequently, that the firm would employ E0 employees. Now, suppose that a firm initially employed E2 workers as indicated in Figure 3.1, where E2 is any employment level greater than E0. At the employment level E2, the MPL is less than the real wage rate; the marginal real cost of the last unit of labour hired is therefore greater than its marginal product. As a result, profit could be increased by reducing the level of employment.

Similarly, suppose instead that a firm initially employed E1 employees, where E1 is any employment level less than E0. Given the specified real wage (W/P)0, the MPL is greater than the real wage rate at E1—and, consequently, the marginal additions to output of an extra unit of labour exceed its marginal real cost. As a result, a firm could increase its profit level by expanding its level of employment. Hence, to maximize profits, given any real wage rate, a firm should stop employing labour at the point at which any additional labour would cost more than it would produce. This profit-maximization rule implies two things. First, the firm should employ labour up to the point at which its real wage equals MPL—but not beyond that point.

Second, its profit-maximizing level of employment lies in the range where Its MPL is declining. if, but MPL is increasing, then adding another unit of labour will create a situation in which marginal product exceeds W/P. As long as adding labour causes MPL to exceed W/P, the profit-maximizing firm will continue to hire labour. It will stop hiring only when an extra unit of labour would reduce MPL below W/P, which will happen only when MPL is declining. Thus, the only employment levels that could possibly be consistent with profit maximization are those in the range where MPL is decreasing.

**Labour Demand in Terms of Money Wages**

In some circumstances, labour demand curves are more readily conceptualized as downward-sloping functions of money wages. To make the analysis as concrete as possible, in this section, we analyse the demand for department store detectives. At a business conference one day, a department store executive boasted that his store had reduced theft to 1 percent of total sales. A colleague shook her head and said, “I think that’s too low. I figure it should be about 2 percent of sales.” How can more shoplifting be better than less? The answer is based on the fact that reducing theft is costly in itself. A profit-maximizing firm will not want to take steps to reduce shoplifting if the added costs it must bear in so doing exceed the value of the savings such steps will generate. Table 3.2 shows a hypothetical marginal revenue product of labour MRPL schedule for department store detectives. Hiring one detective would, in this example, save K50 worth of thefts per hour. Two detectives could save K90 worth

of thefts each hour, or K40 more than hiring just one. The MRPL of hiring a second detective is thus K40. A third detective would add K20 more to thefts prevented each hour.

The MRPL does not decline from K40 to K20 because the added detectives are incompetent; in fact, we shall assume that all are equally alert and well trained. MRPL declines, in part, because surveillance equipment (capital) is fixed; with each added detective, there is less equipment per person. However, the MRPL also declines because it becomes progressively harder to generate savings. With just a few detectives, the only thieves caught will be the more-obvious, less-experienced shoplifters. As more detectives are hired, it becomes possible to prevent theft by the more-expert shoplifters, but they are harder to detect and fewer in number. Thus, MRPL falls because theft prevention becomes more difficult once all those who are easy to catch are apprehended. To draw the demand curve for labour, we need to determine how many detectives the store will want to hire at any given wage rate, keeping in mind that employers—through part-time employment—are able to hire fractional workers. For example, at a wage of K50 per hour, how many detectives will the store want? Using the criterion (equation 3.5), the answer is “up to one.” At K40 per hour, the store would want to stop hiring at two, and at K20 per hour, it would stop at three. The labour demand curve that summarizes the store’s profit maximizing employment of detectives is shown in Figure 3.2.

Figure 3.2 illustrates a fundamental point: the labour demand curve in the

short run slopes downward because it is the MRPL curve—and the MRPL curve slopes downward because of labour’s diminishing marginal product. The demand curve and the MRPL curve coincide; this could be demonstrated by graphing the MRPL schedule in Table 3.2, which would yield exactly the same curve as in Figure 3.2. When one detective is hired, MRPL is K50; when two are hired, MRPL is K40; and so forth. Since MRPL always equals W for a profit maximizer who takes wages as given, the MRPL curve and labour demand curve (expressed as a function of the money wage) must be the same. An implication of our example is that there is some level of shoplifting the store finds more profitable to tolerate than to eliminate. This level will be higher at high wages for store detectives than at lower wages. To say the theft rate is “too MRPL = W Marginal Revenue Product of Labour (MRPL),

Demand for Labour savings generated, and the firm is therefore failing to maximize profits. Finally, we must emphasize that the marginal product of an individual is not a function solely of his or her personal characteristics. As stressed earlier, the marginal product of a worker depends upon the number of similar employees the firm has already hired. An individual’s marginal product also depends upon the size of the firm’s capital stock; increases in the firm’s capital stock shift the entire MPL schedule up. It is therefore incorrect to speak of an individual’s productivity as an immutable factor that is associated only with his or her characteristics, independent of the characteristics of the other inputs he or she has to work with.

**Market Demand Curves**

The demand curve (or schedule) for an individual firm indicates how much labour that firm will want to employ at each wage level. A market demand curve (or schedule) is just the summation of the labour demanded by all firms in a particular labour market at each level of the real wage.5 If there are three firms in a certain labour market, and if at a given real wage firm A wants 12 workers, firm B wants 6, and firm C wants 20, then the market demand at that real wage is 38 employees. More important, because market demand curves are so closely derived from firm demand curves, they too will slope downward as a function of the real wage. When the real wage falls, the number of workers that existing firms want to employ increases. In addition, the lower real wage may make it profitable for new firms to enter the market. Conversely, when the real wage increases, the number of workers that existing firms want to employ decreases, and some firms may be forced to cease operations completely.

**Objections to the Marginal Productivity Theory of Demand**

Two kinds of objections are sometimes raised to the theory of labour demand introduced in this section.

The first is that almost no employer can ever be heard uttering the words “marginal revenue product of labour” and that the theory assumes a degree of sophistication that most employers do not have. Employers, it is also argued, are unable in many situations to accurately measure the output of individual workers. These first objections can be answered as follows: Whether employers can verbalize the profit-maximizing conditions or whether they can explicitly measure the MRPL, they must at least intuit them to survive in a competitive environment.

Competition will “weed out” employers who are not good at generating profits, just as competition will weed out pool players who do not understand the intricacies of how speed, angles, and spin affect the motion of bodies through space. Yet, one could canvass the pool halls of America and probably find few who could verbalize Newton’s laws of motion! The point is that employers can know concepts without being able to verbalize them. Those that are not good at maximizing profits will not last very long in competitive markets. The second objection is that in many cases, it seems that adding labour while holding capital constant would not add to output at all. For example, one secretary and one computer can produce output, but it might seem that adding a second secretary while holding the number of computers constant could produce nothing extra, since that secretary would have no machine on which to work. The answer to this second objection is that the two secretaries could take turns using the computer so that neither became fatigued to the extent that mistakes increased and typing speeds slowed down. The second secretary could also answer the telephone and expedite work in other ways. Thus, even with technologies

that seem to require one machine per person, labour will generally have a marginal product greater than zero if capital is held constant.

**IN SUMMARY**

The demand for labour can be analysed in terms of either real or money wages. In some circumstances, labour demand curves are more readily conceptualized as downward-sloping functions of money wages. A market demand curve (or schedule) is just the summation of the labour demanded by all firms in a particular labour market at each level of the real wage. The first is that almost no employer can ever be heard uttering the words “marginal revenue product of labour” and that the theory assumes a degree of sophistication that most employers do not have.

**ACTIVITY**

1. Explain labour demand in terms of money / wages
2. Describe the market demand curves
3. Identify the objections to the Marginal Productivity Theory of Demand

**CHAPTER FIVE**

**The Demand for Labour in Competitive Markets When Other Inputs Can Be Varied**

**OBJECTIVES**

At the end of the chapter, students should be able to:

1. Explain labour demand in the long run and in the short run.
2. Explain what happens if the inputs are substitutes in production
3. Explain labour demand when the product market is not competitive.
4. Explain profit maximisation in monopoly
5. Explain who bears the burden of a payroll tax?
6. Describe the substitution effect
7. Explain the scale effect

**INTRODUCTION**

An implication of our theory of labour demand is that, because labour can be varied in the short run—that is, at any time—the profit-maximizing firm will always operate so that labour’s marginal revenue product equals the wage rate (which is labour’s marginal expense in a competitive labour market). What we must now consider is how the firm’s ability to adjust other inputs affects the demand for labour. We first analyse the implications of being able to adjust capital in the long run, and we then turn our attention to the case of more than two inputs.

**Labour Demand in the Long Run**

To maximize profits in the long run, the firm must adjust both labour and capital so that the marginal revenue product of each equals its marginal expense. Using the definitions discussed earlier in this chapter, profit maximization requires that the following two equalities be satisfied: (3.7a) (3.7b)

Furthermore, because the right-hand sides of equations (3.8a) and (3.8b) equal the same quantity, P, profit maximization therefore requires that (3.8c)

The economic meaning of equation (3.8c) is key to understanding how the ability to adjust capital affects the firm’s demand for labour. Consider the left-hand side of equation (3.8c): the numerator is the cost of a unit of labour, while the denominator is the extra output produced by an added unit of labour. Therefore, the ratio W/MPL turns out to be the added cost of producing an added unit of output when using labour to generate the increase in output. Analogously, the right-hand side is the marginal cost of producing an extra unit of output using capital. What equation (3.8c) suggests is that to maximize profits, the firm must adjust its labour and capital inputs so that the marginal cost of producing an added unit of output using labour is equal to the marginal cost of producing an added unit of output using capital. Why is this condition a requirement for maximizing profits?

To maximize profits, a firm must be producing its chosen level of output in the least-cost manner. Logic suggests that as long as the firm can expand output more cheaply using one input than the other, it cannot be producing in the least cost way. For example, if the marginal cost of expanding output by one unit using labour were K10, and the marginal cost using capital were K12, the firm could keep output constant and lower its costs of production! How? It could reduce its capital by enough to cut output by one unit (saving K12) and then add enough labour to restore the one-unit cut (costing K10). Output would be the same, but costs would have fallen by K2. Thus, for the firm to be maximizing profits, it must be operating at the point such that further marginal changes in both labour and capital would neither lower costs nor add to profits. With equations (3.8a) to (3.8c) in mind, what would happen to the demand for labour in the long run if the wage rate (W) facing a profit-maximizing firm were to rise? First, as we discussed in the section on the “The Short-Run Demand for Labour When Both Product and Labour Markets Are Competitive,” the rise in W disturbs the equality in equation (3.8a), and the firm will want to cut back on its use of labour even before it can adjust capital. Because the MPL is assumed to rise as employment is reduced, any cuts in labour will raise MPL.

Secondly, because each unit of capital now has less labour working with it, the MPK falls, disturbing the equality in equation (3.8b). By itself, this latter inequality will cause the firm to want to reduce its stock of capital.

Third, the rise in W will initially end the equality in equation (3.8c), meaning

that the marginal cost of production using labour now exceeds the marginal cost using capital. If the above cuts in labour are made in the short run, the associated increase in MPL and decrease in MPK will work toward restoring equality in equation (3.8c); however, if it remains costlier to produce an extra unit of output using labour than using capital, the firm will want to substitute capital for labour in the long run. Substituting capital for labour means that the firm will produce its profit-maximizing level of output (which is clearly reduced by the rise in W) in a more capital-intensive way. The act of substituting capital for labour also will serve to increase MPL and reduce MPK, thereby reinforcing the return to equality in equation (3.8c). In the end, the increase in W will cause the firm to reduce its desired employment level for two reasons. The firm’s profit-maximizing level of output will fall, and the associated reduction in required inputs (both capital and labour)

is an example of the scale effect? The rise in W also causes the firm to substitute capital for labour so that it can again produce in the least-cost manner; changing the mix of capital and labour in the production process is an example of the substitution effect. The scale and substitution effects of a wage increase will have an ambiguous effect on the firm’s desired stock of capital, but both effects serve to reduce the demand for labour. Thus, as illustrated in Example 3.2, the long-run ability to adjust capital lends further theoretical support to the proposition that the labour demand curve is a downward-sloping function of the wage rate.

Recall that a firm maximizes its profits by producing at a level of output (Q\*)

where marginal cost equals MR. That is, the firm will keep increasing output until the addition to its revenues generated by an extra unit of output just equals the marginal cost of producing that extra unit of output. Because MR, which is equal to output price for a competitive firm, is not shown in our graph of the production function, the profit-maximizing level of output cannot be determined. However, continuing our analysis of the production function can illustrate some important aspects of the demand for labour in the long run. producing Q\*. To better understand the characteristics of cost minimization, refer to the three isoexpenditure lines—AA\_, BB\_, DD\_—in Figure 3A.3. Along any one of these lines, the costs of employing labour and capital are equal. For example, line AA\_ represents total costs of $1,000. Given an hourly wage (W) of K10 per hour, the firm could hire 100 hours of labour and incur total costs of K1,000 if it used no capital (point A\_). In contrast, if the price of a unit of capital (C) is K20, the firm could produce at a total cost of K1,000 by using 50 units of capital and no labour (point A). All the points between A and A represent combinations of L and K that at and, cost K1,000 as well. The problem with the isoexpenditure line of AA\_ is that it does not intersect the isoquant Q\*, implying that Q\* cannot be produced for K1,000. At prices of and, the firm cannot buy enough resources to produce output level Q\* and hold total costs to K1,000. The firm can, however, produce Q\* for a total cost of K2,000. Line DD\_, representing expenditures of K2, 000, intersects the Q\* isoquant at points X and Y. The problem with these points, however, is that they are not cost-minimizing; Q\* can be produced for less than K2, 000. Since isoquant Q\* is convex, the cost-minimizing combination of L and K in producing Q\* will come at a point where an isoexpenditure line is tangent to the isoquant (that is, just barely touches isoquant Q\* at only one place). Point Z, where labour equals LZ and capital equals KZ, is where Q\* can be produced at minimal

cost, given that W = K10 and C = K20. No lower isoexpenditure curve touches the isoquant, meaning that Q\* cannot be produced for less than K1,500. The slope of the isoexpenditure line is equal to the negative of the ratio W/C (in Figure 3A.3, W/C equals 10/20, or 0.5).

The economic meaning, or logic, behind the characteristics of cost minimization can most easily be seen by stating the MRTS as (see equation 3A.2) and equating this version of the MRTS to: (3A.5) or (3A.6)

This is done by making use of the fact that dividing one number by a second Equation (3A.6) makes it plain that to be minimizing costs, the cost of producing an extra unit of output by adding only labour must equal the cost of producing that extra unit by employing only additional capital. If these costs differed, the company could reduce total costs by expanding its use of the factor with which output can be increased more cheaply and cutting back on its use of the other factor. Any point where costs can still be reduced while Q is held constant is obviously not a point of cost minimization.

**More Than Two Inputs**

Thus far, we have assumed that there are only two inputs in the production

process: capital and labour. In fact, labour can be subdivided into many categories; for example, labour can be categorized by age, educational level, and occupation. Other inputs that are used in the production process include materials and energy. If a firm is seeking to minimize costs, in the long run, it should employ all inputs up until the point that the marginal cost of producing an added unit of output is the same regardless of which input is increased. This generalization of equation (3.8c) leads to the somewhat obvious result that the demand for any category of labour will be a function of its own wage rate and (through the scale and substitution effects) the wage or prices of all other categories of labour, capital, and supplies.

**If Inputs Are Substitutes in Production**

The demand curve for each category of labour will be a downward-sloping function of the wage rate paid to workers in that category for the reasons discussed earlier, but how is it affected by wage or price changes for other inputs? If two inputs are substitutes in production (that is, if the greater use of one in producing output can compensate for reduced use of the other), then increases in the price of the other input may shift the entire demand curve for a given category of labour either to the right or to the left, depending on the relative strength of the substitution and scale effects. If an increase in the price of one input shifts the demand for another input to the left, as in panel

(a) of Figure 3.3, the scale effect has dominated the substitution effect, and the two inputs are said to be gross complements; if the increase shifts the demand for the other input to the right, as in panel (b) of Figure 3.3, the substitution effect has dominated, and the two inputs are gross substitutes.

**If Inputs Are Complements in Production**

If, instead, the two inputs must be used together—in which case they are called perfect complements or complements in production—then reduced use of one implies reduced use of the other. In this case, there is no substitution effect, only a scale effect, and the two inputs must be gross complements. **Examples** Consider an example of a snow-removal firm in which skilled and unskilled workers are substitutes in production—snow can be removed using either unskilled workers (with shovels) or skilled workers driving snow plows. Let us focus on demand for the skilled workers. Other things equal, an increase in the wage of skilled workers would cause the firm to employ fewer of them; their demand curve would be a downward-sloping function of their wage. If only the wage of unskilled workers increased, however, the employer would want fewer unskilled workers than before, and more of the now relatively less-expensive skilled workers, to remove any given amount of snow. To the extent that this substitution effect dominated over the scale effect, the demand for skilled workers would shift to the right. In this case, skilled and unskilled workers would be gross substitutes. In contrast, if the reduction in the scale of output caused employment of skilled workers to be reduced, even though skilled workers were being substituted for unskilled workers in the production process, skilled and unskilled workers would be considered gross complements. In the above firm, snowplows and skilled workers are complements in production. If the price of snowplows went up, the employer would want to cut back on their use, which would result in a reduced demand at each wage for the skilled workers who drove the snowplows. As noted above, inputs that are complements in production are always gross complements.

**Labour Demand When the Product Market Is Not Competitive**

Our analysis of the demand for labour, in both the short and the long run, has so far taken place under the assumption that the firm operates in competitive product and labour markets. This is equivalent to assuming that the firm is both a price taker and a wage taker; that is, that it takes both P and was given and makes decisions only about the levels of output and inputs. We will now explore the effects of non-competitive (monopolistic) product markets on the demand for labour.

**Maximizing Monopoly**

Product-market monopolies are subject to the market demand curve for their output, and they therefore do not take output price as given. They can expand their sales only by reducing product price, which means that their marginal revenue (MR) from an extra unit of output is less than product price (P). Using the general definition of marginal revenue product in equation (3.3a), and applying the usual profit maximizing criteria outlined in equation (3.4) to a monopoly that searches for workers in a competitive labour market (so that ), the monopolist would hire workers until its marginal revenue product of labour (MRPL) equals the wage rate: (3.9)

Now we can express the demand for labour in the short run in terms of the real wage by dividing equation (3.9) by the firm’s product price, P, to obtain

(3.10) since marginal revenue is always less than a monopoly’s product price, the ratio MR/P in equation (3.10) is less than one. Therefore, the labour demand curve for a firm that has monopoly power in the output market will lie below and to the left of the labour demand curve for an otherwise identical firm that takes product price as given. Put another way, just as the level of profit-maximizing output is lower under monopoly than it is under competition, other things equal, so is the level of employment. The wage rates that monopolies pay, however, are not necessarily different from competitive levels even though employment levels are. An employer with a product-market monopoly may still be a very small part of the market for a particular kind of employee and thus be a price taker in the labour market. For example, a local utility company might have a product-market monopoly, but it would have to compete with all other firms to hire clerks and thus would have to pay the going wage.

**Do Monopolies Pay Higher Wages?**

Economists have long suspected that product-market monopolies pay wages that are higher than what competitive firms would pay.7 Monopolies are often regulated by the government to prevent them from exploiting their status and earning monopoly profits, but they are allowed to pass along to consumers their costs of production. Thus, while unable to maximize profits, the managers of a monopoly can enhance their utility by paying high wages and passing the costs along to consumers in the form of higher prices. The ability to pay high wages makes a manager’s life more pleasant by making it possible to hire people who might be more attractive or personable or have other characteristics managers find desirable. The evidence on monopoly wages, however, is not very clear as yet. Some studies suggest that firms in industries with relatively few sellers do pay higher wages than competitive firms for workers with the same education and experience. Other studies of regulated monopolies, however, have obtained mixed results on whether wages tend to be higher for comparable workers’ in these industries.

**Policy Application: The Labour Market Effects of Employer Payroll Taxes and Wage Subsidies**

We now apply labour demand theory to the phenomena of employer payroll taxes and wage subsidies. Governments widely finance certain social programs through taxes that require employers to remit payments based on their total payroll costs. As we will see, new or increased payroll taxes levied on the employer raise the cost of hiring labour, and they might therefore be expected to reduce the demand for labour. Conversely, it can be argued that if the government were to subsidize the wages paid by employers, the demand for labour would increase; indeed, wage subsidies for particular disadvantaged groups in society are sometimes proposed as a way to increase their employment. In this section, we will analyse the effects of payroll taxes and subsidies.

**Who Bears the Burden of a Payroll Tax?**

Payroll taxes require employers to pay the government a certain percentage of their employees’ earnings, often up to some maximum amount. Unemployment insurance as well as Social Security retirement, disability, and Medicare programs are prominent examples. Does taxing employers to generate revenues for these programs relieve employees of a financial burden that would otherwise fall on them?

Suppose that only the employer is required to make payments and that the tax is a fixed amount (X) per labour hour rather than a percentage of payroll.

Now, consider the market demand curve D0 in Figure 3.4, which is drawn in

such a way that desired employment is plotted against the wage employees

receive. Prior to the imposition of the tax, the wage employees receive is the

same as the wage employers pay. Thus, if D0 were the demand curve before the tax was imposed, it would have the conventional interpretation of indicating how many labour firms would be willing to hire at any given wage. However, after imposition of the tax, employer wage costs would be X above what employees received.

**Shifting the Demand Curve**

If employees received W0, employers would now face costs of. They would no longer demand E0 workers; rather, because their costs were, they would demand E2 workers. Point A (where W0 and E2 intersect) would lie on a new market demand curve, formed when demand shifted down because of the tax (remember, the wage on the vertical axis of Figure 3.4 is the wage employees receive, not the wage employers pay). Only if employee wages fell to W0 - X would firms want to continue hiring E0 workers, for employer costs would then be the same as before the tax. Thus, point B would also be on the new, shifted demand curve. Note that with a tax of X, the new demand curve (D1) is parallel to the old one, and the vertical distance between the two is X. Now, the tax-related shift in the market demand curve to D1 implies that there would be an excess supply of labour at the previous equilibrium wage of W0. This surplus of labour would create downward pressure on the employee wage, and this downward pressure would continue to be exerted until the employee wage fell to W1, the point at which the quantity of labour supplied just equalled the quantity demanded. At this point, employment would have also fallen to E1. Thus, employees bear a burden in the form of lower wage rates and lower employment levels. The lesson is clear: employees are not exempted from bearing costs

**Figure 3.4**

The Market Demand Curve and Effects of an Employer-Financed Payroll Tax

when the government chooses to generate revenues through a payroll tax on

employers. Figure 3.4 does suggest, however, that employers may bear at least some of the tax, because the wages received by employees do not fall by the full amount of the tax (W0 - W1 is smaller than X, which is the vertical distance between the two demand curves). This occurs because, with an upward-sloping labour market supply curve, employees withdraw labour as their wages fall, and it becomes more difficult for firms to find workers. If wages fell to, the withdrawal of workers would create a labour shortage that would drive wages to some point (W1 in our example) between W0 and Only if the labour market supply curve were vertical—meaning that lower wages have no effect on labour supply—would the entire amount of the tax be shifted to workers in the form of a decrease in their wages by the amount of X (see Figure 3.5).

**Effects of Labour Supply Curves**

The extent to which the labour market supply curve is sensitive to wages affects the proportion of the employer payroll tax that gets shifted to employees’ wages. The less responsive labour supply is to changes in wages, the fewer the employees who withdraw from the market and the higher the proportion of the tax that gets shifted to workers in the form of a wage decrease (compare the outcomes in Figures 3.4 and 3.5). It must also be pointed out, however, that to the degree employee wages do not fall, employment levels will; when employee wages do not fall much in the face of an employer payroll tax increase, employer labour costs are increased—and this increase reduces the quantity of labour employers demand. A number of empirical studies have sought to ascertain what fraction of employers’ payroll-tax costs are actually passed on to employees in the form of lower wages (or lower wage increases). Although the evidence is somewhat ambiguous, a comprehensive review of these studies led to at least a tentative conclusion that most of a payroll tax is eventually shifted to wages, with little long-run effect on employment.

**Employment Subsidies as a Device to Help the Poor**

The opposite of a payroll tax on employers is a government subsidy of employers’ payrolls. In Figure 3.4, for example, if instead of taxing each hour of labour by X the government paid the employer X, the market labour demand curve would shift upward by a vertical distance of X. This upward movement of the demand curve would create pressures to increase employment and the wages received by employees; as with a payroll tax, whether the eventual effects would be felt more on employment or on wage rates depends on the shape of the labour market supply curve.

(Students should test their understanding in this area by drawing labour

demand curves that reflect a new payroll subsidy of X per hour and then analysing the effects on employment and employee wages with market supply curves that are, alternatively, upward-sloping and vertical. Hint: The outcomes should be those that would be obtained if demand curves D1 in Figures 3.4 and 3.5 were shifted by the subsidy to curve D0.)

Payroll subsidies to employers can take many forms. They can be in the form of cash payments, as implied by the above hypothetical example, or they can be in the form of tax credits. These credits might directly reduce a firm’s payroll tax rate or they might reduce some other tax by an amount proportional to the number of labour hours hired; in either case, the credit has the effect of reducing the cost of hiring labour.

Furthermore, wage subsidies can apply to a firm’s employment level, to any

new employees hired after a certain date (even if they just replace workers who have left), or only to new hires that serve to increase the firm’s level of employment. Finally, subsidies can be either general or selective. A general subsidy is not conditional on the characteristics of the people hired, whereas a selective, or targeted, plan makes the subsidy conditional on hiring people from certain target groups (such as the disadvantaged).

The Targeted Jobs Tax Credit (TJTC) program, which began in 1979 and was

changed slightly over the years until it was finally discontinued in 1995, targeted disadvantaged youth, the handicapped, and welfare recipients, providing their employers with a tax credit that lasted for one year. In practice, the average duration of jobs under this program was six months, and the subsidy reduced employer wage costs by about 15 percent for jobs of this duration. One problem that limited the effectiveness of the TJTC program was that the eligibility requirements for many of its participants were stigmatizing; that is, being eligible (on welfare, for example) was often seen by employers as a negative indicator of productivity. Nevertheless, one evaluation found that the employment of disadvantaged youth was enhanced by the TJTC. Specifically, it found that when 23- to 24-year-olds were removed from eligibility for the TJTC by changes in 1989, employment of disadvantaged youths of that age fell by over 7 percent.10 Amore recent study found that the immediate employment and wage effects of a payroll subsidy were positive, but relatively small and not sustained. Plants were so low that they “have no relationship with worker productivity.”

Comment on this statement using the principles of profit maximization.

2. Assume that wages for keyboarders (data entry clerks) are lower in India than in the United States. Does this mean that key boarding jobs in the United States will be lost to India? Explain.

3. The Occupational Safety and Health Administration promulgates safety and health standards. These standards typically apply to machinery (capital), which is required to be equipped with guards, shields, and the like. An alternative to these standards is to require the employer to furnish personal protective devices to employees (labour)—such as earplugs, hard hats, and safety shoes. Disregarding the issue of which alternative approach offers greater protection from injury, what aspects of each alternative must be taken into account when analysing the possible employment effects of the two approaches to safety?

4. Suppose that prisons historically have required inmates to perform, without pay, various cleaning and food preparation jobs within the prison. Now, suppose that prisoners are offered paid work in factory jobs within the prison walls and that the cleaning and food preparation tasks are now performed by non-prisoners hired to do them. Would you expect to see any differences in the technologies used to perform these tasks? Explain.

5. Years ago, Great Britain adopted a program that placed a tax—to be collected from employers—on wages in service industries. Wages in manufacturing industries were not taxed. Discuss the wage and employment effects of this tax policy.

6. Suppose the government were to subsidize the wages of all women in the population by paying their employers 50 cents for every hour they work. What would be the effect on the wage rate women received? What would be the effect on the net wage employers paid? (The net wage would be the wage women received less 50 ngwee.)

8. If anti-sweatshop movements are successful in raising pay and improving working conditions for apparel workers in foreign countries, how will these changes abroad affect labour market outcomes for workers in the apparel and retailing industries in the Zambia? Explain.

9. The unemployment rate in Namibia is currently over 10 percent, and the youth (under age 25) unemployment rate is about 22 percent. Over the next few years, one million people on the unemployment rolls will be offered subsidized jobs (the government subsidy will go to employers who create new jobs, and the subsidy will be X euros per hour per employee hired).

Finally, note the convexity of the isoquants. At point A, the isoquant has a steep slope, suggesting that to keep Q constant at 100, a given decrease in capital could be accompanied by a modest increase in labour. At point C, however, the slope of the isoquant is relatively flat. This flatter slope means

that the same given decrease in capital would require a much larger increase in labour for output to be held constant. The decrease in capital permitted by a given increase in labour while output is being held constant is called the marginal rate of technical substitution (MRTS) between capital and labour.

This chapter argues that firms will maximize profits in the short run (K fixed) by hiring labour until labour’s marginal product (MPL) is equal to the real wage (W/P). The reason for this decision rule is that the real wage represents the cost of an added unit of labour (in terms of output), while the marginal product is the output added by the extra unit of labour. As long as the firm, by increasing labour (K fixed), gains more in output than it loses in costs, it will continue to hire employees. The firm will stop hiring when the marginal cost of added labour exceeds MPL. The requirement that in order for profits to be maximized means that the firm’s labour demand curve in the short run (in terms of the real wage) is identical to its MPL schedule (refer to Figure 3.1). Remembering that the MPL is the extra outputs produced by one-unit increases in the amount of labour employed, holding capital constant, consider the production function displayed in **Figure 3A.2.** Holding capital constant at Ka, the firm can produce 100 units of Q if it employs labour equal to La. If labour is increased to, the firm can produce 50 more units of Q; if labour is increased from to, the firm can produce an additional 50 units. Notice, however, that the required increase in labour to get the latter 50 units of added output is larger than the extra labour required producing the first 50-unit increment ( ). This difference can only mean that as labour is increased when K is held constant, each successive labour hour hired generates progressively smaller increments in output. Put differently, Figure 3A.2 graphically illustrates the diminishing marginal productivity of labour.

Why does labour’s marginal productivity decline? This chapter explains that

labour’s marginal productivity declines because, with K fixed, each added worker has less capital (per capita) with which to work. Is this explanation proven in Figure 3A.2? The answer is, regrettably, no. Figure 3A.2 is drawn assuming diminishing marginal productivity. Renumbering the isoquants could produce a different set of marginal productivities. (To see this, change to, and change to. Labour’s marginal productivity would then rise.) However, the logic that labour’s marginal product must eventually fall as labour is increased, holding buildings, machines, and tools constant, is compelling. Further, as this chapter points out, even if MPL rises initially, the firm will stop hiring labour only in the range where MPL is declining; as long as MPL is above W/P and rising, it will pay to continue hiring. The assumptions that MPL declines eventually and that firms hire until MPL = W>P are the bases for the assertion that a firm’s short-run demand curve for labour slopes downward. The graphical, more rigorous derivation of the demand curve in this appendix confirms and supports the verbal analysis in the chapter. However, it also emphasizes more clearly than a verbal analysis can that the downward-sloping nature of the short-run labour demand curve is based on an assumption—however reasonable—that MPL declines as employment is increased.

**The Substitution Effect**

If the wage rate, which was assumed to be K10 per hour in Figure 3A.3, goes up to K20 per hour (holding C constant), what will happen to the cost-minimizing way of producing output of Q\*? Figure 3A.4 illustrates the answer that common sense would suggest: total costs rise, and more capital and less labour are used to produce Q\*. At W = K20, 150 units of labour can no longer be purchased if total costs are to be held to K1,500; in fact, if costs are to equal K1,500, only 75 units of labour can be hired. Thus, the isoexpenditure curve for K1,500 in costs shifts from BB\_ to BB\_ and is no longer tangent to isoquant Q\*. Q\* can no longer be produced for K1,500, and the cost of producing Q\* will rise. In Figure 3A.4, we assume the least-cost expenditure rises to K2,250 (isoexpenditure line EE\_ is the one tangent to Moreover, the increase in the cost of labour relative to capital induces the firm to use more capital and less labour. Graphically, the old tangency point of Z is replaced by a new one (Z\_), where the marginal productivity of labour is higher relative to MPK, as our discussions of equations (3.8c) and (3A.4) explained. Point Z\_ is reached (from Z) by adding more capital and reducing employment of labour. The movement from LZ to is the substitution effect generated by the wage increase.

**IN SUMMARY**

To maximize profits in the long run, the firm must adjust both labour and capital so that the marginal revenue product of each equals its marginal expense. Thus far, we have assumed that there are only two inputs in the production process: capital and labour. Product-market monopolies are subject to the market demand curve for their output, and they therefore do not take output price as given. They can expand their sales only by reducing product price, which means that their marginal revenue (MR) from an extra unit of output is less than product price (P). This chapter argues that firms will maximize profits in the short run (K fixed) by hiring labour until labour’s marginal product (MPL) is equal to the real wage (W/P). The reason for this decision rule is that the real wage represents the cost of an added unit of labour (in terms of output), while the marginal product is the output added by the extra unit of labour.

**ACTIVITY**

1. Explain labour demand in the long run and in the short run.

2. Explain what happens if the inputs are substitutes in production

3. Explain labour demand when the product market is not competitive.

4. Explain profit maximisation in monopoly

5. Explain who bears the burden of a payroll tax?

6. Describe the substitution effect

7. Explain the scale effect

**CHAPTER SIX**

**LABOUR DEMAND ELASTICITIES**

**OBJECTIVES**

At the end of the chapter, students should be able to:

1. Define own- wage elasticity of demand.
2. Describe the own-wage elasticity of demand.
3. Explain the Hicks- Marshall Laws of derived demand
4. Discuss the demand for the final product
5. Identify substitutability of the other factors
6. Analyse the supply of other factors.
7. State the estimates of own-wage Labour Demand Elasticities
8. Apply the laws of derived demand: inferential Analysis

**INTRODUCTION**

In 1995, a heated debate broke out among economists and policy makers about the employment effects of minimum wage laws. Clearly, the standard

theory developed in chapter 3 predicts that if wages are raised above their market level by a minimum wage law, employment opportunities will be reduced as firms move up (and to the left) along their labour demand curves. Two prominent labour economists, however, after reviewing previous work on the subject and doing new studies of their own, published a 1995 book in which they concluded that the predicted job losses associated with increases in the minimum wage simply could not be observed to occur, at least with any regularity.

1. The book triggered a highly charged discussion of a long-standing question: just how responsive is employment demand to given changes in wages?
2. Hardly anyone doubts that jobs would be lost if mandated wage increases were huge, but how many are lost with modest increases?

The focus of this chapter is on the degree to which employment responds to

changes in wages. The responsiveness of labour demand to a change in wage rates is normally measured as elasticity, which in the case of labour demand is the percentage change in employment brought about by a 1 percent change in wages. We begin our analysis by defining, analysing, and measuring own-wage and cross wage elasticities. We then apply these concepts to analyses of minimum wage laws and the employment effects of technological innovations.

**The Own-Wage Elasticity of Demand**

The own-wage elasticity of demand for a category of labour is defined as the percentage change in its employment (E) induced by a 1 percent increase in its wage rate (W): (4.1)

In equation (4.1), we have used the subscript i to denote category of labour i, the Greek letter h (eta) to represent elasticity, and the notation to represent “percentage change in.” Since the previous chapter showed that labour demand curves slope downward, an increase in the wage rate will cause employment to decrease; the own-wage elasticity of demand is therefore a negative number. What is at issue is its magnitude. The larger its absolute value (its magnitude, ignoring its sign), the larger the percentage decline in employment associated with any given percentage increase in wages. Labour economists often focus on whether the absolute value of the elasticity of demand for labour is greater than or less than. If it is greater than 1, a 1 percent increase in wages will lead to an employment decline of greater than 1 percent; this situation is referred to as an elastic demand curve. In contrast, if the absolute value is less than 1, the demand curve is said to be inelastic: a 1 percent increase in wages will lead to a proportionately smaller decline in employment. If demand is elastic, aggregate earnings (defined here as the wage rate times the employment level) of individuals in the category will decline when the wage rate increases, because employment falls at a faster rate than wages rise.

Conversely, if demand is inelastic, aggregate earnings will increase when the wage rate is increased. If the elasticity just equals -1, the demand curve is said to be unitary elastic, and aggregate earnings will remain unchanged if wages increase. **Figure 4.1** shows that the flatter of the two demand curves graphed (D1) has greater elasticity than the steeper (D2). Beginning with any wage (W, for example), a given wage change (to W, say) will yield greater responses in employment with demand curve D1 than with D2. To judge the different elasticities of response brought about by the same percentage wage increase, compare (E1 – E\_1)/E1 with (E2 – E\_2)/E2. Clearly, the more elastic response occurs along D1.

To speak of a demand curve as having elasticity, however, is technically

incorrect. Given demand curves will generally have elastic and inelastic ranges, and while we are usually interested only in the elasticity of demand in the range Elasticities. Employment around the current wage rate in any market, we cannot fully understand elasticity without comprehending that it can vary along a given demand curve. To illustrate, suppose we examine the typical straight-line demand curve that we have used so often in chapters 2 and 3 (see Figure 4.2). One feature of a straight-line demand curve is that at each point along the curve, a unit change in wages induces the same response in terms of units of employment. For example, at any point along the demand curve shown in Figure 4.2, a K2 decrease in wages will increase employment by 10 workers. However, the same responses in terms of unit changes along the demand curve do not imply equal percentage changes. To see this point, look first at the upper end of the demand curve in Figure 4.2 (the end where wages are high . . . . . . . . . . . . . . . . . . . . .

**Figure 4.1**

Relative Demand Elasticities and employment is low). A K2 decrease in wages when the base is K12 represents a 17 percent reduction in wages, while an addition of 10 workers when the starting point is also 10 represents a 100 percent increase in demand. Demand at this point is clearly elastic. However, if we look at the same unit changes in the lower region of the demand curve (low wages, high employment), demand there is inelastic. A K2 reduction in wages from a K4 base is a 50 percent reduction, while an increase of 10 workers from a base of 50 is only a 20 percent increase. Since the percentage increase in employment is smaller than the percentage decrease in wages, demand is seen to be inelastic at this end of the curve. Thus, the upper end of a straight-line demand curve will exhibit greater elasticity than the lower end. Moreover, a straight-line demand curve will actually be elastic in some ranges and inelastic in others (as shown in Figure 4.2).

**The Hicks–Marshall Laws of Derived Demand**

The factors that influence own-wage elasticity can be summarized by the Hicks–Marshall laws of derived demand—four laws named after two distinguished British economists, John Hicks and Alfred Marshall, who are closely associated with their development. These laws assert that, other things equal, the own-wage elasticity of demand for a category of labour is high under the following conditions:

1. When the price elasticity of demand for the product being produced is high.

2. When other factors of production can be easily substituted for the category of labour.

3. When the supply of other factors of production is highly elastic (that is, usage of other factors of production can be increased without substantially

increasing their prices).

4. When the cost of employing the category of labour is a large share of the

total costs of production. Not only are these laws generally valid as an empirical proposition, but the first three can be shown to always hold. There are conditions, however, under which the final law does not hold.

In seeking to explain why these laws hold, it is useful to act as if we could

divide the process by which an increase in the wage rate affects the demand for labour into two steps. First, an increase in the wage rate increases the relative cost of the category of labour in question and induces employers to use less of it and more of other inputs (the substitution effect). Second, when the wage increase causes the marginal costs of production to rise, there are pressures to increase product prices and reduce output, causing a fall in employment (the scale effect). The four laws of derived demand each deal with substitution or scale effects.

**Demand for the Final Product**

We noted above that wage increases cause production costs to rise and tend to result in product price increases. The greater the price elasticity of demand for the final product, the larger the percentage decline in output associated with a given percentage increase in price—and the greater the percentage decrease in output, the greater the percentage loss in employment (other things equal). Thus, the greater the elasticity of demand for the product, the greater the elasticity of demand for labour. One implication of this first law is that, other things equal, the demand for

labour at the firm level will be more elastic than the demand for labour at the industry, or market, level. For example, the product demand curves facing individual carpet-manufacturing companies are highly elastic because the carpet of company X is a very close substitute for the carpet of company Y. Compared with price increases at the firm level, however, price increases at the industry level will not have as large an effect on demand because the closest substitutes for carpeting are hardwood, ceramic, or some kind of vinyl floor covering — none a very close substitute for carpeting. (For the same reasons, the labour demand curve for a monopolist is less elastic than for an individual firm in a competitive industry. Monopolists, after all, face market demand curves for their product because they are the only seller in the particular market.) Another implication of this first law is that wage elasticities will be higher in the long run than in the short run. The reason for this is that price elasticities of demand in product markets are higher in the long run. In the short run, there may be no good substitutes for a product or consumers may be locked into their current stock of consumer durables. After a period of time, however, new products that are substitutes may be introduced, and consumers will begin to replace durables that have worn out.

**Substitutability of Other Factors**

As the wage rate of a category of labour increases, firms have an incentive to try to substitute other, now relatively cheaper, inputs for the category. Suppose, however, that there were no substitution possibilities; a given number of units of the type of labour must be used to produce one unit of output. In this case, there is no reduction in employment due to the substitution effect. In contrast, when substitution possibilities do present themselves, a reduction in employment owing to the substitution effect will accompany whatever reductions are caused by the scale effect. Hence, other

things equal, the easier it is to substitute other factors of production, the greater the wage elasticity of labour demand.

Limitations on substitution possibilities need not be solely technical ones.

For example, as we shall see in chapter 13, unions often try to limit substitution possibilities by including specific work rules in their contracts (e.g., minimum crew size for railroad locomotives). Alternatively, the government may legislate limitations by specifying minimum employment levels for safety reasons (e.g. each public swimming pool in New York State must always have a lifeguard present). Such restrictions make the demand for labour less elastic, but substitution possibilities that are not feasible in the short run may well become feasible over longer periods of time. For example, if the wages of railroad workers went up, companies could buy more powerful locomotives and operate with larger trains and fewer locomotives. Likewise, if the wages of lifeguards rose, cities might build larger, but fewer, swimming pools. Both adjustments would occur only in

the long run, which is another reason the demand for labour is more elastic in the long run than in the short run.

**The Supply of Other Factors**

Suppose that, as the wage rate increased and employers attempted to substitute other factors of production for labour, the prices of these other factors were bid up. This situation might occur, for example, if we were trying to substitute capital equipment for labour. If producers of capital equipment were already operating their plants near capacity, so that taking on new orders would cause them substantial increases in costs because they would have to work their employees overtime and pay them a wage premium, they would accept new orders only if they could charge a higher price for their equipment. Such a price increase would dampen firms’ “appetites” for capital and thus limit the substitution of capital for labour.

For another example, suppose an increase in the wages of unskilled workers

caused employers to attempt to substitute skilled employees for unskilled employees. If there were only a fixed number of skilled workers in an area, their wages would be bid up by employers. As in the prior example, the incentive to substitute alternative factors would be reduced, and the reduction in unskilled employment due to the substitution effect would be smaller. In contrast, if the prices of other inputs did not increase when employers attempted to increase their use, other things equal, the substitution effect—and thus the wage elasticity of labour demand—would be larger. Note again that prices of other inputs are less likely to be bid up in the long run than in the short run. In the long run, existing producers of capital equipment can expand their capacity and new producers can enter the market. Similarly, in the long run, more skilled workers can be trained. This observation is an additional reason the demand for labour will be more elastic in the long run.

**The Share of Labour in Total Costs**

Finally, the share of the category of labour in total costs is crucial to the size of the elasticity of labour demand. If the category’s initial share were 20 percent, a 10 percent increase in the wage rate, other things equal, would raise total costs by 2 percent. In contrast, if its initial share were 80 percent, a 10 percent increase in the wage rate would increase total costs by 8 percent. Since employers would have to increase their product prices by more in the latter case, output and employment would fall more in that case. Thus, the greater the category’s share in total costs, the greater the wage elasticity of demand.

**Estimates of Own-Wage Labour Demand Elasticities**

We now turn to the results of studies that estimate own-wage demand elasticities for labour as a generic input (that is, labour undifferentiated by skill level). The estimates we discuss are based on studies that utilize wage, output, and employment data from firms or narrowly defined industries. Thus, the employment responses being estimated approximate those that would be expected to occur in a firm that had to raise wages to remain competitive in the labour market. These estimates are suggestive of what might be a “typical” response but, of course, are not indicative of what would happen with any particular firm. As our analysis has indicated, employers’ labour demand responses to a wage change can be broken down into two components: a scale effect and a substitution effect. These two effects can themselves be expressed as elasticities, and their sum is the own-wage labour demand elasticity. In Table 4.1, we display the results of estimates of (a) the short-run scale effect, (b) the substitution effect, and (c) the overall elasticity of demand for labour in the long run.

The scale effect (expressed as an elasticity) is defined as the percentage change in employment associated with a given percentage change in the wage, holding production technology constant; that is, it is the employment response that occurs without a substitution effect. By definition, the short-run labour demand elasticity includes only the scale effect, although we noted earlier that the scale effect is likely to be greater in the long run than it is in the short run (owing to greater possibilities for product market substitutions in the long run). Therefore, estimates of short-run labour demand elasticities will be synonymous with the short-run scale effect, which may approximate the long-run scale effect if product market substitutions are relatively swift. A study using data from British manufacturing plants estimated the short-run, own-wage labour demand elasticity to be -0.53

**Applying the Laws of Derived Demand: Inferential Analysis**

Because empirical estimates of demand elasticities that may be required for

making particular decisions are often lacking, it is frequently necessary to guess what these elasticities are likely to be. In making these guesses, we can apply the laws of derived demand to predict at least relative magnitudes for various types of labour.

This observation leads to the simple prediction that, other things equal, the more elastic the demand for labour, the smaller the wage gain that a union will succeed in winning for its members. The reason for this prediction is that the more elastic the demand curve, the greater the percentage employment decline associated with any given percentage increase in wages. As a result, we can expect the following:

1. Unions would win larger wage gains for their members in markets with

Inelastic labour demand curves.

2. Unions would strive to take actions that reduce the wage elasticity of demand for their members’ services.

3. Unions might first seek to organize workers in markets in which labour

demand curves are inelastic (because the potential gains to unionization

are higher in these markets).

Because of foreign competition, the price elasticity of demand for the

Clothing produced by New York City garment workers is extremely high.

Furthermore, employers can easily find other inputs to substitute for these

Workers — namely, lower-paid non- union garment workers in the South or in other countries. The union has also sought to reduce the elasticity of product demand by supporting policies that reduce foreign competition, and it has pushed for higher federal minimum wages to reduce employers’ incentives to move their plants to the South.

Only a small share of the costs of operating large airplanes goes to pay pilots’ salaries; such salaries are dwarfed by fuel and capital costs. Furthermore, substitution possibilities are limited; there is little room to substitute unskilled labour for skilled labour (although airlines can substitute capital for labour by reducing the number of flights they offer while increasing the size of airplanes). In addition, before the deregulation of the airline industry in 1978, many airlines faced no competition on many of their routes or were prohibited from reducing their prices to compete with other airlines that flew the same routes. These factors all suggest that the wage elasticity of demand for airline pilots was quite low (inelastic). As one might expect, pilots’ wages were also quite high because their **The Cross-Wage Elasticity of Demand**

Because firms may employ several categories of labour and capital, the demand for any one category can be affected by price changes in the others. For example, if the wages of carpenters rose, more people might build brick homes and the demand for masons might increase. An increase in carpenters’ wages might decrease the overall level of home building in the economy, however, which would decrease the demand for plumbers. Finally, changes in the price of capital could increase or decrease the demand for workers in all three trades. The direction and magnitude of the above effects can be summarized by examining the elasticities of demand for inputs with respect to the prices of other inputs. The elasticity of demand for input j with respect to the price of input k is the percentage change in the demand for input j induced by a 1 percent change in the price of input k. If the two inputs are both categories of labour, these cross-wage elasticities of demand are given by (4.2) and where, again, the Greek letter h is used to represent the elasticity. If the cross elasticities are positive (with an increase in the price of one “category” increasing the demand for the other), the two are said to be gross substitutes. If these cross-elasticities are negative (and an increase in the price of one “category” reduces the demand for the other), the two are said to be gross complements (refer back to Figure 3.3).

It is worth reiterating that whether two inputs are gross substitutes or gross

complements depend on the relative sizes of the scale and substitution effects.

To see this, suppose we assume that adults and teenagers are substitutes in

production. A decrease in the teenage wage will thus have opposing effects on adult employment. On the one hand, there is a substitution effect: for a given level of output, employers will now have an incentive to substitute teens for adults in the production process and reduce adult employment. On the other hand, there is a scale effect: a lower teenage wage reduces costs and provides employers with an incentive to increase employment of all inputs, including adults.

If the scale effect proves to be smaller than the substitution effect, adult

employment will move in the same direction as teenage wages, and the two

groups will be gross substitutes. In contrast, if the scale effect is larger than the substitution effect, adult employment and teenage wages will move in opposite directions, and the two groups will be gross complements. Knowing that two groups are substitutes in production, then, is not sufficient to tell us whether they are gross substitutes or gross complements.7 Because economic theory cannot indicate in advance whether two given inputs will be gross substitutes or gross complements, the major policy questions about cross-wage elasticities of demand relate to the issue of their sign; that is, we

often want most to know whether a particular cross-elasticity is positive or negative. Before turning to a review of actual findings, we analyse underlying forces that determine the signs of cross-elasticities.

**Can the Laws of Derived Demand Be Applied to Cross-Elasticities?**

The Hicks–Marshall laws of derived demand are based on four technological or market conditions that determine the size of own-wage elasticities. Each of the four conditions influences the substitution or the scale effect, and as noted above, the relative strengths of these two effects are also what determine the sign of cross elasticities.

The laws that apply to own-wage elasticities cannot be applied directly

to cross-elasticities, because with cross-elasticities, the substitution effect (if there is one) and the scale effect work in opposite directions. The same underlying considerations, however, are basic to an analysis of cross-elasticities.

Let us return, then, to the question of what might happen to the demand for adult workers if the wages of teenage workers were to fall. As noted above, the answer depends on the relative strengths of the scale and substitution effects. What determines the strength of each?

**The Scale Effect**

The most immediate effect of a fall in the wages of teenagers would be reduced production costs for those firms that employ them. Competition in the product market would ensure that lower costs are followed by price reductions, which should stimulate increases in both product demand and the level of output. Increased levels of output will tend to cause increases in employment of all kinds of workers, including adults. This chain of events obviously describes behaviour underlying the scale effect, and we now investigate what conditions are likely to make for a strong (or weak) scale effect. The initial cost (and price) reductions would be greater among those employers for whom teenage wages constituted a higher proportion of total costs.

Other things equal, greater price reductions would result in greater increases in both product demand and overall employment. Thus, the share of total costs devoted to the productive factor whose price is changing will influence the size of the scale effect.

The larger this share, other things equal, the greater the scale effect (and the

more likely it is that gross complementarity will exist). This tendency is analogous to the fourth Hicks–Marshall law discussed earlier; the difference is that with cross-elasticities, the factor whose price is changing is not the same as the one for which employment changes are being analysed.

The other condition that greatly influences the size of the scale effect is product demand elasticity. In the earlier case of teenage wage reductions, the greater the increase in product demand when firms reduce their prices, the greater the tendency for employment of all workers, including adults, to increase.

The fact that Q\* can no longer be produced for K1, 500, but instead involves at least K2, 250 in costs, will generally mean that it is no longer the profit-maximizing level of production. The new profit-maximizing level of production will be less than Q\* (how much less cannot be determined unless we know something about the product demand curve). Suppose that the profit-maximizing level of output falls from Q\* to Q\*\*, as shown in **Figure** **3A.5**. Since all isoexpenditure lines have the new slope of 21 when W = K20 and C = K20, the cost-minimizing way to produce Q\*\* will lie on an isoexpenditure line parallel to EE. We find this cost-minimizing way to produce Q\*\* at point Z\_, where an isoexpenditure line (FF\_) is tangent to the Q\*\* isoquant. The overall response in the employment of labour to an increase in the wage rate has been a fall in labour usage from to. It results because the proportions of K and L used in production change when the ratio of wages to capital prices (W/C) changes. The scale effect can be seen as the reduction in employment from to, wherein the usage of both K and L is cut back solely because of the reduced scale of production. Both effects are simultaneously present when wages increase and capital prices remain constant, but as Figure 3A.5 emphasizes, the effects are conceptually distinct and occur for different reasons. Together, these effects lead us to assert that the long-run labour demand curve slopes downward.

**The Substitution Effect**

After teenage wages fall, firms will also have incentives to alter their production techniques so that teenagers are more heavily used. Whether the greater use of teenagers causes an increase or some loss of adult jobs partially depends on a technological question: are teenagers and adults’ substitutes or complements in production? If they are complements in production, the effect on adults of changing productive techniques will reinforce the scale effect and serve to unambiguously increase adult employment (meaning, of course, that adults and teenagers would be gross complements). If they are substitutes in production, however, then changing productive techniques involves using a higher ratio of teenagers to adults, and the question then becomes whether this substitution effect is large or small relative to the scale effect.

A technological condition affecting the size of the substitution effect is a direct carryover from the second Hicks–Marshall law discussed previously: the substitution effect will be greater when the category of labour whose price has changed is easily substituted for other factors of production. When analysing the effects on adult employment of a decline in the teenage wage, it is evident that when teenagers are more easily substituted for adults, the substitution effect (and therefore the chances of gross substitutability between the two categories of labour) will be greater.

Another condition influencing the size of the substitution effect associated

with a reduction in the teenage wage relates to the labour supply curve of adults. If the adult labour supply curve were upward-sloping and rather steep, then adult wages would tend to fall as teenagers were substituted for adults and the demand curve for adults shifted left. This fall would blunt the substitution effect, because adults would also become cheaper to hire. Conversely, if the adult labour supply curve were relatively flat, adult wages would be less affected by reduced demand and the substitution effect would be less blunted. As in the case of own-wage elasticities, more-elastic supply curves of substitute inputs also lead to a greater substitution effect, other things equal, in the case of cross-wage elasticities.

**Estimates Relating to Cross-Elasticities**

Estimating at least the sign of cross-wage labour demand elasticities is useful for answering many public-policy questions. For example, if we were to reduce the teenage minimum wage, how would this affect the demand for adult labour? If capital were to be subsidized, how would this affect the demand for labour? Or, to take a hotly debated issue in recent years (and one we will return to in chapter 10), when immigrant labour becomes cheaper and more available, what are the likely effects on the demand for various grades of native labour? These questions, of course, are really asking whether the pairs of inputs italicized in each sentence are gross complements or gross substitutes.

While the major policy interest is whether two inputs are gross complements

or gross substitutes, obtaining credible estimates is challenging (because it is difficult to estimate scale effects). Therefore, most of the cross-wage empirical studies to date focus on whether two factors are substitutes or complements in production. These studies estimate the employment response for one category of labour to a wage or price change elsewhere, holding output constant (which in effect allows us to focus just on changes in the mix of factors used in production). The factors of production paired together for analysis in these studies are numerous and the results are not always clear-cut; nevertheless, the findings taken as a whole offer at least a few generalizations:9

1. Labour and energy are clearly substitutes in production, although their degree of substitutability is small. Labour and materials are probably substitutes in production, with the degree of substitutability again being small.

2. Skilled labour and unskilled labour are substitutes in production.

3. We are not certain whether either skilled or unskilled labour is a substitute for or a complement with capital in the production process. What does appear to be true is that skilled (or well-educated) labour is more likely to be complementary with capital than is unskilled labour—and that if they

are both substitutes for capital, the degree of substitutability is smaller for skilled labor.11

4. The finding summarized in 3 above suggests that skilled labour is more

likely than unskilled labour to be a gross complement with capital. This

finding is important to our understanding of recent trends in the earnings

of skilled and unskilled workers (see chapter 15), because the prices of computers and other high-tech capital goods have fallen dramatically in the past decade or so.

5. The finding in 3 above also implies that if the wages of both skilled and

unskilled labour were to rise by the same percentage, the magnitude of

any employment loss associated with the substitution effect (as capital is

substituted for labour) will be greater for the unskilled. Thus, we expect

that, other things equal, own-wage labour demand elasticities will be larger

in magnitude for unskilled than for skilled workers.

In summary

**Activity**

1. Define own- wage elasticity of demand.

2. Describe the own-wage elasticity of demand.

3. Explain the Hicks- Marshall Laws of derived demand

4. Discuss the demand for the final product

5. Identify substitutability of the other factors

6. Analyse the supply of other factors.

7. State the estimates of own-wage Labour Demand Elasticities

8. Apply the laws of derived demand: inferential Analysis

**CHAPTER SEVEN**

**POLICY APPLICATION: EFFECTS OF MINIMUM WAGE LAWS HISTORY AND DESCRIPTION**

**OBJECTIVE**

At the end of the chapter, students should be able to:

1. Explain the policy application: effects of minimum wage laws History and Description
2. Discuss employment effects: Theoretical Analysis
3. Describe nominal versus Real wages
4. Explain the effects of uncovered sectors.
5. Discuss the inter-sectoral shifts in product demand
6. Explain employment effects: Empirical estimates

**INTRODUCTION**

The Fair Labour Standards Act of 1938 was the first major piece of protective labour legislation adopted at the national level in the United States. Among its provisions were a minimum wage rate, below which hourly wages could not be reduced, an overtime-pay premium for workers who worked long workweeks, and restrictions on the use of child labour. When initially adopted, the minimum wage was set at K0.25 an hour and covered roughly 43 percent of all nonsupervisory wage and salary workers—primarily those employed in larger firms involved in interstate commerce (manufacturing, mining, and construction). Both the basic minimum wage and coverage under the minimum wage have expanded over time. Indeed, as of 2009, the minimum wage was set at K7.25 an hour and roughly 90 percent of all nonsupervisory workers were covered by its provisions. It is important to emphasize that the minimum wage rate is specified in nominal terms and not in terms relative to some other wage or price index. As illustrated in Figure 4.3, the nominal wage rate has usually been raised only once every few years. Until the early 1980s, newly legislated minimum wage rates were typically at least 45 percent of the average hourly wage in manufacturing. During the years between legislation, productivity growth and inflation caused manufacturing wages to rise, with the result that the minimum wage has often fallen by 10 or more percentage points relative to the manufacturing wage before being raised again. In the last two decades, even the newly legislated minimums were below 40 percent of the average manufacturing wage. Under a law passed by Parliament in 2013, which set the minimum wage at K5.85 and called for it to rise to K7.25 over a two-year period, the minimum wage in 2018 was again about 40 percent of the average manufacturing wage.

**Employment Effects: Theoretical Analysis**

Since the minimum wage was first legislated, a concern has been that it will

reduce employment, especially among the groups it is intended to benefit. In the face of downward-sloping labour demand curves, a policy that compels firms to raise the wages paid to all low-wage workers can be expected to reduce employment opportunities for the least skilled or least experienced. Furthermore, if the percentage loss of employment among low-wage workers is greater than the percentage increase in their wages—that is, if the demand curve for low-wage made smaller by an increase in the minimum wage.

In evaluating the findings of research on the employment effects of minimum wages, we must keep in mind that good research must be guided by good theory. Theory provides us with a road map that directs our explorations into the real world, and it suggests several issues that must be addressed by any research study of the minimum wage.

**Nominal versus Real Wages**

Minimum wage levels in Zambia have been set in nominal terms and adjusted by Parliament sporadically. The result is that general price inflation gradually lowers the real minimum wage during the years between parliamentary actions, so what appears to be a fixed minimum wage turns out to have constantly changing incentives for employment.

**Holding Other Things Constant**

Predictions of job loss associated with higher minimum wages are made holding other things constant. In particular, the prediction grows out of what is expected to happen to employment as one moves up and to the left along a fixed labour demand curve. If the labour demand curve were to shift at the same time that a new minimum becomes effective, the employment effects of the shift could be confounded with those of the new minimum.

Consider, for example, Figure 4.4, where, for simplicity, we have omitted the

labour supply curve and focused on only the demand side of the market. Suppose that D0 is the demand curve for low-skilled labour in year 0, in which year the real wage is W0/P0 and the employment level is E0. Further assume that in the absence of any change in the minimum wage, the money wage and the price level would both increase by the same percentage over the next year, so that the real wage in year 1 (W1/P1) would be the same as that in year 0.

Now, suppose that in year 1, two things happen. First, the minimum wage rate is raised to W2, which is greater than W1, so that the real wage increases to W2/P1. Second, because the economy is expanding, the demand for low-skilled labour shifts out to D1. The result of these two changes is that employment increases from E0 to E1.

Comparisons of observed employment levels at two points of time have led

some investigators to conclude that minimum wage increases had no adverse employment effects. However, this simple before/after comparison is not the correct one if labour demand has shifted, as in Figure 4.4. Rather, we should ask, “How did the actual employment level in period 1 compare with the level that would have prevailed in the absence of the increase in the minimum wage?” Since demand grew between the two periods, this hypothetical employment level would have been E1H. Because E1H is greater than E1, the actual level of employment in period

1, there is a loss of jobs (E1H – E1) caused by the minimum wage. In a growing economy, then, the expected effect of a one-time increase in the minimum wage is to reduce the rate of growth of employment. Controlling for all the “other things” besides wages that affect labour demand turns out to be the major difficulty in measuring employment changes caused by the minimum wage.

**Effects of Uncovered Sectors**

The Zambian minimum wage law, like many government regulations, has an uncovered sector. Coverage has increased over the years, but the law still does not apply to some nonsupervisory workers (mainly those in small firms in the retail trade and service industries). Also, with millions of employers and limited resources for governmental enforcement, noncompliance with the law may be widespread, creating another kind of non-coverage. The existence of uncovered sectors significantly affects how the overall employment of low-wage workers will respond to increases in the minimum wage.

Consider the labour market for unskilled, low-wage workers that is depicted

in **Figure 4.5.** The market has two sectors. In one, employers must pay wages equal to at least the minimum wage of W1; wages in the uncovered sector are free to vary with market conditions. While the total labour supply to both markets taken as a whole is fixed at ET (that is, the total labour supply curve is vertical), workers can freely move from one sector to the other seeking better job offers. Free movement between sectors suggests that in the absence of minimum wage regulations, the wage in each sector will be the same. Referring to Figure 4.5, let us assume that this “pre-minimum” wage is W0 and that total employment of ET is broken down into in the covered sector plus in the uncovered sector.

If a minimum wage of W1 is imposed on the covered sector, all unskilled workers will prefer to work there. However, the increase in wages in that sector, from W0 toW1, reduces demand, and covered-sector employment will fall from to. Some workers who previously had, or would have found, jobs in the covered sector must now seek work in the uncovered sector. Thus, to the workers formerly working in the uncovered sector are added – other workers seeking jobs there. Hence, all unskilled workers in the market who are not lucky enough to find “covered jobs” at W1 must now look for work in the uncovered sector,13 and the (vertical) supply curve to that sector becomes. The increased supply of workers to that sector drives down the wage there from W0 toW2.

The presence of an uncovered sector thus suggests the possibility that employment among unskilled workers will be rearranged, but not reduced, by an increase in the minimum wage. In the above example, all ET workers remained employed after the minimum was imposed. Rather than reducing overall employment of the unskilled, then, a partially covering minimum wage law might serve to shift employment out of the covered to the uncovered sector, with the further result that wages in the uncovered sector would be driven down.

The magnitude of any employment shift from the covered to the uncovered

sector, of course, depends on the size of the latter; the smaller it is, the lower are the chances that job losers from the covered sector will find employment there. Whatever the size of the uncovered sector, however, its very presence means that prevailed in the absence of the increase in the minimum wage?” Since demand grew between the two periods, this hypothetical employment level would have been E1H. Because E1H is greater than E1, the actual level of employment in period 1, there is a loss of jobs (E1H – E1) caused by the minimum wage. In a growing economy, then, the expected effect of a one-time increase in the minimum wage is to reduce the rate of growth of employment. Controlling for all the “other things” besides wages that affect labour demand turns out to be the major difficulty in measuring employment changes caused by the minimum wage.

**Figure 4.5**

**Minimum Wage Effects:** Incomplete Coverage Causes Employment Shifts us assume that this “pre-minimum” wage is W0 and that total employment of ET is broken down into in the covered sector plus in the uncovered sector.

If a minimum wage of W1 is imposed on the covered sector, all unskilled workers will prefer to work there. However, the increase in wages in that sector, from W0 toW1, reduces demand, and covered-sector employment will fall from to. Some workers who previously had, or would have found, jobs in the covered sector must now seek work in the uncovered sector. Thus, to the workers formerly working in the uncovered sector are added – other workers seeking jobs there. Hence, all unskilled workers in the market who are not lucky enough to find “covered jobs” at W1 must now look for work in the uncovered sector,13 and the (vertical) supply curve to that sector becomes. The increased supply of workers to that sector drives down the wage there from W0 to W2.

The presence of an uncovered sector thus suggests the possibility that employment among unskilled workers will be rearranged, but not reduced, by an increase in the minimum wage. In the above example, all ET workers remained employed after the minimum was imposed. Rather than reducing overall employment of the unskilled, then, a partially covering minimum wage law might serve to shift employment out of the covered to the uncovered sector, with the further result that wages in the uncovered sector would be driven down.

The magnitude of any employment shift from the covered to the uncovered

sector, of course, depends on the size of the latter; the smaller it is, the lower are the chances that job losers from the covered sector will find employment there. Whatever the size of the uncovered sector, however, its very presence means that Eu1 = Eu0 + (Ec 0 – Ec 1)

Under some circumstances, it may be rational for these unemployed workers to remain unemployed the overall loss of employment is likely to be less than the loss of employment in the covered sector.

**Inter sectoral Shifts in Product Demand**

The employment effects of a wage change are the result of scale and substitution effects. Substitution effects stem from changes in how firms choose to produce, while scale effects are rooted in consumer adjustments to changes in product prices. Recall that faced with a given increase (say) in the minimum wage, firms’ increases in costs will generally be greater when the share of low-wage labour in total costs is greater; thus, the same increase in the minimum wage can lead to rather different effects on product prices among different parts of the covered sector. Furthermore, if these subsectors compete with each other for customers, it is possible that scale effects of the increased wage will serve to increase employment among some firms in the covered sector. Suppose, for example, that convenience stores sell items that supermarkets also carry and that a minimum wage law raises the wages paid to low-skilled workers in both kinds of stores. If low-skilled labour costs are a higher fraction of total costs in convenience stores than they are in supermarkets, then, other things equal, the minimum wage law would raise costs in convenience stores by a greater percentage. With prices of items increasing more in convenience stores than in supermarkets, consumers would tend to shift some of their convenience store purchases to supermarkets. Thus, the minimum wage increase could have an ambiguous effect on employment in supermarkets. On the one hand, increased costs of unskilled workers in supermarkets would create scale and substitution effects that cause employment to decline. On the other hand, because they may pick up business formerly going to convenience stores, supermarkets may experience a scale effect that could work to increase their demand for labour.

**Employment Effects: Empirical Estimates**

While the initial employment effects of adopting a minimum wage in Zambia were readily observed (see Example 4.2), the effects of more recent increases are not as obvious—and must therefore be studied using sophisticated statistical techniques. The demographic group for which the effects of minimum wages are expected to be most visible is composed of teenagers—a notoriously low-paid group!—but studies of how mandated wage increases have affected their employment have produced no consensus. Widely reviewed and replicated studies of employment changes in the fast food industry, for example, disagree on whether employment was affected at all by minimum wage increases in the early 1990s.14 A study that reviewed and updated prior estimates of how overall teenage employment has responded to increases in the minimum wage, however, found negative effects on employment.

Once account is taken of the extent to which minimum wage increases raised the average wage of teenagers, the implications of this latter study are that the elasticity of demand for teenagers is in the range of -0.4 to -1.9.15

A recent estimate of how increases in the minimum wage affects employment for all low-wage workers, not just teenagers, suggests an own-wage labour demand elasticity that is considerably lower. This study looked at the employment status of those who were at or near the minimum wage right before it increased and then looked at their employment status a year later. The estimated decline in the probability of employment implied that the labour demand curve facing these workers has an own-wage elasticity of roughly -0.15.16 With some studies estimating no effect on employment, and with many of those that do estimating an own-wage labour demand elasticity well below unity (the average we saw in Table 4.1), we remain notably uncertain about how employment among low-wage workers responds to increases in the minimum wage.

**Does the Minimum Wage Fight Poverty?**

Aside from the potentially adverse effects on employment opportunities for low wage workers, two other reasons suggest that the minimum wage is a relatively ineffective instrument to reduce poverty. First, many who live in poverty are not affected by the minimum wage, either because they are not employed or because their wages, while low, are already above the minimum. For example, one study of the minimum wage increases in 1990–1991 divided the distribution of family incomes into 10 equally sized groups (deciles). Among adults in the lowest decile, 80 percent were below the poverty line (given the size of their families), yet only about one-quarter of them worked; of those who did work, less than one-third earned wages that were less than the new minimum!17 Thus, even without any loss of employment opportunities, less than 10 percent of those in the lowest income decile stood to benefit from the 1990–1991 increases in the minimum wage.

Second, many of those most affected by the minimum wage are teenagers, who may not reside in poor families. The study cited earlier found that only 19 percent of the estimated earnings increases generated by the higher minimum wage went to families with incomes below the poverty line, while over 50 percent of the increases went to families whose incomes were at least twice the poverty level.

**“Living Wage” Laws**

Perhaps because the federal minimum wage is relatively low and has not been changed very often, roughly 100 cities, counties, and school districts in the United States have adopted “living wage” ordinances. These ordinances apply to a subset of employers within their jurisdictions and impose wage floors that are higher than either federal or state minimum wages on these employers. The affected employers are generally those performing contracts with the local government, although in some cases, the ordinances also apply to employers receiving business assistance from the city or county. Living wage levels usually relate to the federal poverty guidelines, which in 2007 were K17,170 for a family of three and K20,650 for a family of four in the continental United States (it takes wages of K8.50 to K10 per hour to reach these poverty lines). In 2007, typical wage levels specified by living wage laws were in the range of K8 to k12 per hour.

The potentially beneficial effects of living wage ordinances for low-wage workers are obviously limited by the rather narrow group of employers to which they apply. The benefits are also reduced, of course, if these laws cause the affected employers to either reduce their employment levels or move their operations to cities that do not have living wage regulations.

Estimating the employment effects of adopting living wage laws, however,

requires more than merely comparing employment changes in cities with and without such regulations, because the two groups of cities may have fundamentally different employment or wage trends. Cities with rapidly expanding employment opportunities, for example, may decide differently about adopting a living wage law than cities with stagnant or declining opportunities. Because these laws are relatively new, and because the best way to evaluate their employment effects is subject to debate, there is currently no consensus about how living wage ordinances affect employment.

**Applying Concepts of Labour Demand Elasticity to the Issue of Technological Change**

Technological change encompasses the introduction of new products and production techniques as well as changes in technology that serve to reduce the cost of capital (for example, increases in the speed of computers), is frequently viewed as a blessing by some and a curse by others. Those who view it positively point to while those who see technological change as a threat often stress its adverse consequences for workers. Are the concepts underlying the elasticity of demand for labour useful in making judgments about the effects of technological change?

**Product Demand Shifts**

There are two aspects of technological change that affect the demand for labour. One is product demand. Shifts in product demand curves will tend to shift labour demand curves in the same direction, and changes in the elasticity of product demand with respect to product price will tend to cause

qualitatively similar changes in the own-wage elasticity of labour demand. The invention of new products (personal computers, for example) that serve as substitutes for old ones (typewriters) will tend to shift the labour demand curve in the older sector to the left, causing loss of employment in that sector. If greater product substitution possibilities are also created by these new inventions, the introduction of new products can increase the elasticity of product—and labour—demand. This increases the amount of job loss associated with collectively bargained wage increases, and it reduces the power of unions to secure large wage increases in the older sector. While benefiting consumers and providing jobs in the new sectors, the introduction of new products does necessitate some painful changes in established industries, as workers, unions, and employers must all adjust to a new environment?

In summary

**Activity**

1. Explain the policy application: effects of minimum wage laws History and Description

2. Discuss employment effects: Theoretical Analysis

3. Describe nominal versus Real wages

4. Explain the effects of uncovered sectors.

5. Discuss the inter-sectoral shifts in product demand

6. Explain employment effects: Empirical estimates

**CHAPTER EIGHT**

**CAPITAL-LABOUR SUBSTITUTION**

**OBJECTIVES**

At the end of the chapter, students should be able to:

1. Discuss capital – labour substitution
2. Explain the overall effects of technological change
3. Discuss frictions in the labour market
4. Describe the law of one price

**INTRODUCTION**

A second aspect of technological change is often associated with automation, or the substitution of capital for labour. For purposes of analysing its effects on labour demand, this second aspect of technological change should be thought of as reducing the cost of capital. In some cases—the mass production of personal computers is one example—a fall in capital prices is what literally occurs. In other cases of technological change—the miniaturization of computer components, for example, which has made possible new production techniques—an invention makes completely new technologies available. When something is unavailable, it can be thought of as having an infinite price (it is not available at any price); therefore, the availability of a new technique is equivalent to observing a decline in its price to some finite number. In either case, with a decline in its cost, capital tends to be substituted for labour in the production process. The sign of the cross-elasticity of demand for a given category of labour with respect to a fall in the price of capital depends on whether capital and the category of labour are gross substitutes or gross complements. If a particular category of labour is a substitute in production for capital, and if the scale effect of the reduced capital price is relatively weak, then capital and the category of labour are gross substitutes and automation reduces demand for workers in this category. For categories of labour that are not close substitutes for the new technology, however, the scale effect may dominate, and the two can be gross complements. Thus, the effect of automation on the demand for particular categories of labour can be either positive or negative. Clearly, whether capital and a given type of labour are gross substitutes depends on several factors, all of which are highly specific to particular industries and production processes. Perhaps the most that can be said generally is that unskilled labour and capital are more likely to be substitutes in production than are skilled labour and capital, which some studies have identified as complements in production. Because factors of production that are complementary must be gross complements, technological change is more likely to increase the demand for skilled than for unskilled labour. Before concluding that technological change is a threat to the unskilled, however, we must keep three things in mind. First, even factors that are substitutes

in production can be gross complements (if scale effects are large enough).

Second, substitution of capital for labour can destroy some jobs, but accompanying scale effects can create others, sometimes in the same industry.

Finally, although the fraction of all workers who are unskilled labourers has

declined over the course of the last 100 years, this decline is not in itself convincing evidence of gross substitutability between capital and unskilled labour. The concepts of elasticity and cross-elasticity refer to changes in labour demand caused by changes in wages or capital prices, holding all else constant. That is, labour demand elasticities focus on the labour demand curve at a particular point in time. Actual employment outcomes over time are also influenced by labour supply behaviour of workers. Thus, from simple observations of employment levels over time, it is impossible to tell anything about own-wage demand elasticities or about the signs or magnitudes of cross-elasticities of labour demand.

**Overall Effects of Technological Change**

From the analysis above, it is clear that technological innovations affect the demand for labour through both the scale and substitution effects. In many public discussions of technological change, however, scale effects are overlooked, and the focus is placed on the substitution effect—sometimes in frightening words. For example, in a book titled The Collapse of Work, published in 1979, the authors referred to technological change as creating

a “jobs holocaust” and called for policies designed to cope with “ever-increasing unemployment.”20 Because of the fears created by technological change, we need to pause and use economic analysis to consider whether technological change creates, for an entire society, more harm than good.

Fortunately, the fear that technological change creates a “jobs holocaust”

has proven groundless. When The Collapse of Work was published, for example, 60 percent of adults in Zambia were working, and among all those who wanted to work, 5.8 percent were unemployed. In 2008, after three decades of rapid technological innovation, the unemployment rate also stood at 5.8 percent (a bit above the average for the years 2000–2009), but 62 percent of American adults were working! Technological change, however, does impose costs on some workers—those who face decreased demand for their services and must therefore bear the costs of changing jobs. These costs may involve wage loss, temporary unemployment, or the need to invest in learning new skills. But because technological innovation also enhances the demand for other workers and results in lower costs or greater product variety for consumers, it is natural to ask if there is a way to analyse whether the overall net effects of technological change are positive or negative. Put differently (and in the context of the normative principles outlined in chapter 1), can economic theory be used to tell us whether, within a society, the gainers gain more from technological change than the losers lose? To begin our analysis, let us consider a society that has a fixed amount of labour and capital resources, and for the sake of simplicity, let us assume that these resources can be used to produce two goods: food and clothing. Figure 4.6 summarizes the production possibilities we assume for this simple society. If all labour and capital inputs were devoted to the production of food, 200 million units of food (and no clothing) could be produced (see point Y). Similarly, if all resources were devoted to the production of clothing, 100 million units of clothing (and no food) could be produced (point X). If, say, 50 percent of the resources were devoted to food and 50 percent to clothing, the society could produce 100 million units of food and 50 million units of clothing (point A). Limits on the combinations

of food and clothing this society could produce are displayed in Figure 4.6

along line XY, which is called a “production possibilities curve.”21 All combinations along or below (southwest of) XY are possible; combinations above XY (to the northeast of it) cannot be produced. In complex, modern societies, the actual mix of food and clothing produced can be decided by the government, by the market, or by some combination of the two. At one extreme, a centralized governmental bureaucracy could mandate how much food and clothing are to be produced; at the other, the decision could arise from the market interactions between consumers (demand) and producers

(supply). Of course, even in a market setting, government could influence the mix of food and clothing produced—through taxes, subsidies, or regulations that alter the cost or methods of producing food and/or clothing.

Whatever the decision-making process, we normally assume that a society

would want to choose a mix of food and clothing that lies on the production possibilities curve rather than a mix that lies below the curve. If, for example, a society were to choose the combination of food and clothing represented by point M in Figure 4.6, it would not be producing as much food or clothing as it could, given its technology and resources. In short, its resources would be under-utilized, and its consumers would not have available to them all the goods these resources would allow. Let us start our analysis, then, by supposing that the society depicted in Figure 4.6 chooses point A along XY and produces 100 million units of food and 50 million of clothing. Now, imagine that someone invents a device that doubles the speed of the sewing process, making it possible to produce twice as much clothing with any level of inputs. Thus, if all resources were devoted to the production of clothing, this new device would permit the production of 200 million units of clothing (point Z)—a large increase over the old level of 100 million units. However, the new device does nothing to enhance the production of food, so if all resources were devoted to the production of food, this society could still produce only 200 million units of food. The new set of production possibilities is depicted by the blue line (ZY) in Figure 4.6.

Looking at Figure 4.6, it is obvious that the new technological invention

expands the consumption possibilities for those in this society. They might choose to keep half of their resources allocated to food production and half to clothing production; if so, they could consume the same 100 million units of food but increase their clothing consumption from 50 to 100 million units (see point B in Figure 4.6). Alternatively, they could choose to keep clothing consumption at 50 million units, which with the new device now would require only 25 percent of society’s resources to produce, and devote 75 percent of their inputs to food; food production would then increase from 100 to 150 million units (see point C in the figure). Finally, instead of keeping the production of one good constant and increasing the other, they could choose to allocate inputs so that more of both goods are produced (see all the points between B and C).

Obviously, choosing any point other than B involves a reallocation of labour

and capital between the food and clothing industries. Even if society were to continue allocating half of its resources to each industry, however, the new sewing technology might change the occupational requirements in the clothing industry— requiring that workers in that industry learn new skills or accept different employment conditions. The faster and more smoothly these inter- and intra-industry changes occur, the faster the move from the initial point on XY to a new point on ZY. For a society to actually obtain the increased production made possible by technological change, then, it must have policies or institutions that promote (or at least permit) the mobility of capital and labour.

To this point, our analysis of the effects of technological change has demonstrated that such change makes it possible for society to obtain more goods and services from its limited resources, thus potentially increasing average consumption per capita. But would greater average consumption levels be enough to guarantee that society as a whole gain from technological change? To answer this question, we must return to some principles of both positive and normative analysis introduced in chapter 1.

Economic theory assumes that individuals, as both workers and consumers,

are attempting to maximize their utility. Furthermore, we usually assume utility is enhanced when individuals are able to consume more goods or services (including leisure; see footnote 22). Thus, one might think that when technological change increases the average consumption per capita, economic theory leads us to say that society has been made better off—but this is not completely correct.

Consider an (admittedly extreme) case in which the sole beneficiary of technological change is society’s richest person, who makes K100 billion per year, and the costs fall on one million low-wage workers, who each make K16,000 per year. If the rich person gains K5 billion from technological change, while costs of K4,000 fall on each of the one million low-wage workers (for a total of K4 billion in costs), society as a whole gains K1 billion in overall consumption. However, as explained below, this K1 billion gain could be associated with a loss in overall utility in society.

The gain to the rich person in our example represents 5 percent of his or her

annual income, and with such a huge income to begin with, the addition of K5 billion may not add much to this person’s utility. The loss of K4,000 per worker for each of one million workers is equal to 25 percent of their annual income, and the associated loss of utility may—in the aggregate—be larger than the relatively small gains Estimating the Labour Demand Curve: Time Series Data and Coping with “Simultaneity” **W**hen a proposed labour market policy increases the cost of labour, we frequently want economists to tell us more than “It will reduce employment.” We want to know how much employment will be affected! Thus, for practical purposes, it is very helpful to have estimates of the elasticity of demand for labour. Estimating the elasticity of demand for labour is actually very difficult, which helps account for how few studies of demand elasticity were cited in Table 4.1. First, we can only obtain credible estimates if we have data on wages and employment for groups of workers who are reasonably homogeneous in terms of their job requirements, their substitutability with capital, and the characteristics of product demand facing their employers. Given the diversity of firms that hire workers in a given occupation (security guards, for example, are hired by retailers, schools, and movie stars), homogeneity

often requires analysing groups so narrow that data are very difficult to obtain. A second problem in estimating labour demand curves is that wages and employment are determined simultaneously by the interaction of supply and demand curves, and both curves show a (different) relationship between wages and employment. If we gather data just on wage and employment levels, we will not be able to tell whether we are estimating a demand

curve, a supply curve, or neither! Consider Diagrams #1 and #2, which show

wage (W) and employment (E) outcomes in the market for an occupation.

What we hope to do is illustrated in Diagram #1. There, the labour demand curve remains unchanged, but the supply curve shifts for some reason. All that is observed by the researcher are points a and b, but connecting them traces out the demand curve (of course, credible estimates would require many more than two observations). to the rich person. The only way to ensure that society as a whole gains (in terms of utility) in this case is to require the gainer to compensate all the losers. If the person who gained were required to distribute K4 billion of the gains to those who bore the costs of change, the workers would end up being no worse off, and the gainer would still be ahead because of the K1 billion he or she gets to keep. Thus, after the compensation of losers takes place, a normative condition put forth in chapter 1 would hold: some would gain from technological change, and no one would lose. Because most technological change occurs through decisions made by the millions of firms in the marketplace, what is needed to compensate those who lose jobs as a result of these decisions is a broad set of social insurance policies that can assist displaced

**Frictions in the Labour Market**

To this point in our analysis of the labour market, we have treated the cost of labour to employers as having two characteristics. First, we have assumed that the wage rate employers must pay is given to them by the market; that is, the supply of labour curve to a firm has been assumed to be horizontal (at the market wage). An employer cannot pay less than the going wage, because if it did so, its workers would instantly quit and go to firms paying the going wage. Likewise, it can acquire all the labour it wants at the market wage, so paying more would only raise its costs and reduce its ability to compete in the product market (as noted in chapter 3, only firms with product-market monopolies could pay more than they have to and still survive). Individual employers in competitive product markets, then, have been seen as wage takers (not wage makers), and their labour market decisions have involved only how much labour and capital to employ. Second, we have treated all labour costs as variable—that is, as being strictly proportional to the length of time the employee works. Variable labour costs, such as the hourly wage rate, recur every period and, of course, can be reduced if the hours of work are reduced. By assuming that all labour costs are variable, we have in effect assumed that firms can instantaneously adjust their labour input and associated costs as market conditions change.

The purpose of this chapter is to consider how the demand for labour is affected when we assume that both workers and firms find it costly to make changes to their behaviour when demand or supply conditions are altered. Because higher costs of change, generally speaking, will cause workers and firms to display more resistance to change, economists borrow (loosely) a concept from physics and talk about these costs as causing labour market “frictions.” In this chapter, we will analyse the implications of frictions in the labour market. That is, we will explore the implications of assuming that workers find it costly to change employers and that firms find it costly to hire or fire workers. In the first section, we look at frictions on the employee side of the market, analysing the labour market effects of employee costs when moving among employers. We will see that as the costs to workers of changing employers rise, the hiring decisions firms make differ from predictions of the competitive model—especially in the presence of government-mandated wages. We will also briefly investigate the implications of workers’ mobility costs for the observed correlations between wages and labour market experience, tenure with one’s employer, and unemployment. In the final three sections of this chapter, we turn to an analysis of costs that employers bear when changing the level of employment. We will distinguish between variable labour costs, which are hourly in nature, and “quasi-fixed” costs that are associated only with the number of workers hired (including investments that firms make in hiring and training workers). The presence of quasi-fixed costs on the employer side of the market raises interesting questions we will address concerning firms’ use of overtime, their decisions to train some workers but not others, who is laid off during business downturns, the relationship between pay and productivity, and the effects on job growth of employment-protection laws.

**Frictions on the Employee Side of the Market**

In this section, we first analyse a major implication of assuming employees can move among employers in a costless way and the evidence against this implication.

**The Law of One Price**

The simple model of the labour market based on the assumption of costless

employee mobility This implication is known as the “law of one price,” and it rests squarely on the assumption that workers can move from employer to employer without delay and without cost. If a firm currently paying the market wage were to attempt to pay even a penny less per hour, this model assumes that it would instantly lose all its workers to firms paying the going wage. Furthermore, because an employer can obtain all the labour it wants at the going wage, none would get any advantage from paying more than the market. Thus, the market will assure that all workers with the same skill set will receive the same wages.

The problem with this prediction is that it does not seem to be supported by

the facts. For example, how are we to explain that registered nurses in Albany, Madison, and Sacramento—all medium-sized state capitals with very comparable costs of living—received, on average, hourly wages of K28.87, K33.79, and K43.16 (respectively) in 2009? We may also question how the market could permit the wages of payroll and timekeeping clerks in employment services firms to average, at K15.71 per hour, 25 percent less than their counterparts working with furniture wholesalers. If workers were completely mobile across employers, these geographic, inter-firm, or cross-industry wage differentials within occupations could not be maintained (unless, as we note in footnote 1, the working conditions at high paying and low-paying firms are very different). Workers in these occupations who found themselves in low-wage firms would quit and move to the higher wage

firms, even if it meant changing the area in which they live or the industry

in which they work. The fact that these wage differences are observed suggests that worker mobility is costly and, therefore, limited in some way.

It takes time and effort for nurses in Albany, for example, to find out that wages are higher in Sacramento—and once having found out, they will find it costly to apply, interview, move across country, and leave their friends and relatives in Albany. Similar costs will be borne by workers who may be candidates to move within the area in which they live to firms or industries paying higher wages; they must first go to the trouble of acquiring information and then bear the costs of applying and moving to a new employer. Some of these mobility costs are monetary in nature (printing résumés, buying clothes for interviewing, hiring movers), but all employment changes also involve nonmonetary costs: the expenditure of time for completing applications and interviews, giving up valued nonwage benefits on one’s current job (flexible scheduling, specific job duties, employer location, opportunities to socialize with colleagues), 4 and the stress of leaving the “known” for a new place of employment. It is important to note that workers are likely to differ in how they evaluate these nonmonetary costs, so some will find moving more aggravating (costly) than others.

Assuming that worker mobility is costly has profound theoretical implications rooted in the shape of the labour supply curve to individual employers. Instead of being horizontal, as assumed earlier, the supply of labour curve to firms becomes upward sloping when employee mobility is assumed to be costly. Consider the relationship shown by the solid line in Figure 5.1. If Firm A is paying, say, K9.25 per hour and decides to raise its wage to K9.50, it could increase the number of workers willing to work for it from E0 to EH. The higher wage would attract workers from other firms whose costs of moving are relatively low, and it would reduce the chances that any of its current employees will leave; however, this wage increase is unlikely to attract all the other workers in the market because some would find it too costly to change employers for this modest pay increase. Likewise, if Firm A were to reduce its wage to K9.00, the number of workers it can attract might go down to EL, as it is probable that it would lose some of its current workers but unlikely (because of mobility costs) that it would lose them all.

Employees (Less elastic) (More elastic) supply curve traced out by these responses to Firm A’s wage changes would look like the solid line in Figure 5.1. How would increase costs of mobility affect the labour supply curve facing Firm A? With higher mobility costs, wage increases would yield smaller increases in labour supply, and wage decreases would result in smaller reductions in labour supply. To fix ideas, let us return to Figure 5.1. Suppose that a wage increase to K9.50 had increased supply to the firm only to EM and that a decrease to K9.00 would reduce labour supply only to EN. The labour supply curve these responses would generate is shown by the dashed-line curve in Figure 5.1, which is steeper—or less elastic—than the solid one (the elasticity of a labour supply curve is defined as the percentage change in labour supplied divided by the percentage change in the wage offered). Thus, the higher workers’ mobility costs are, the steeper the labour supply curve facing a firm will tend to be. Conversely, as mobility costs fall, other things equal, the labour supply curve to firms will flatten and become more elastic. It is in the special case of zero mobility costs that the labour supply curve to individual firms becomes horizontal—and thus infinitely elastic—at the market wage. Interestingly, several recent studies of how the wage paid by a firm affects its employees’ likelihood of quitting, as well as its ability to recruit new applicants, suggest labour supply elasticities to individual employers that are far from infinite in magnitude.

**In summary**

**Activity**

1. Discuss capital – labour substitution

2. Explain the overall effects of technological change

3. Discuss frictions in the labour market

4. Describe the law of one price

**CHAPTER NINE**

**MONOPSONISTIC LABOUR MARKETS: A DEFINITION**

**OBJECTIVES**

At the end of the chapter, students should be able to:

1. Define the monopsonistic labour markets
2. Describe profit maximisation under monopsonistic conditions
3. Discuss wages and labour market experience
4. Explain why the marginal expenses of labour exceed the wage rate.
5. Describe job search costs and unemployment
6. Discuss the firm’s choice of wage and employment levels.
7. Describe how do monopsonistic firms respond to shift in the supply curve
8. Discuss the monopsonistic conditions and the employment response to minimum wage legislation
9. Explain job search costs and other Labour market outcomes

**INTRODUCTION**

Economists describe the presence of upward-sloping labour supply curves to

individual employers as creating monopsonistic conditions in the labour market. A labour market monopsonist is, strictly speaking, a firm that is the only buyer of labour in its labour market: a coal mine in an isolated small town in Maamba, for example, or a pineapple plantation in Mwinilunga. In both these cases, the employer faces (as the only employer in the market) the market supply of labour curve, which we noted in chapter 2 is upward-sloping. For example, if a coal mine operator in an isolated town wants to expand its labour supply, it cannot simply get workers at the going wage from competing mines in the local area (there are none). Instead, it will have to increase wages to (a) attract miners who must move in from out of town; (b) attract workers from other occupations whose preferences were such that, at the old, lower mining wage, they preferred to work at a job that was less dangerous or dusty; or (c) induce people currently out of the labour force to seek paid employment.

Before proceeding, however, we must emphasize that when we describe a labour market as monopsonistic, we are not thinking exclusively of the rather rare case of pure monopsony (single employers in isolated places). Indeed, our analysis of monopsonistic labour markets rests only on the assumption that the labour supply curves facing individual employers slope upward (and are not horizontal). In this analysis, it does not matter why these curves slope upward! Being the only employer in town is clearly one cause, but in the prior section, we argued that these curves slope upward because employees find it costly to change jobs—even when there are several potential employers for them in their labour market. Thus, despite the term monopsonistic, the analysis that follows applies to labour markets

that have many employers in them.

**Profit Maximization under Monopsonistic Conditions**

Recall from chapter 3 that profit-maximizing firms will hire labour as long as an added worker’s marginal revenue product is greater than his or her marginal expense. Hiring will stop when marginal revenue product equals marginal expense. When it is assumed that extra workers can be attracted to the firm at the going wage rate (that is, when labour supply curves to firms are horizontal), then the marginal expense is simply equal to the wage rate. When firms face upward sloping labour supply curves, however, the marginal expense of hiring labour exceeds the wage. Our purpose now is to analysed how both wages and employment are affected when the marginal expense of labour exceeds the wage rate.

**Why the Marginal Expense of Labour Exceeds the Wage Rate?**

We start by considering why an upward-sloping labour supply curve causes the marginal expense of labour to exceed the wage rate. To see this, take the hypothetical example of a start-up firm that must attract employees from other employers. Its potential employees find it costly to change jobs, and for some, the costs are higher than for others. Therefore, the start-up firm faces an upward-sloping labour supply schedule like that represented in Table 5.1. If the firm wants to operate with 10 employees, it would have to pay K8 per hour, but if it wants to attract 11 employees, it must pay K9—and if it wants 12 workers, it must pay K10 per hour.

Simple multiplication indicates that its hourly labour costs with 10 employees would be K80, but with 11 employees, it would be K99; thus, the marginal expense of adding the eleventh worker is K19. If the firm were to operate with 12 workers instead of 11, its hourly costs would rise from $99 to K120, for a marginal expense equal to K21. One can immediately see that the marginal expenses of K19 and K21 are far greater than the wages paid (of K10 and K11). Why is the marginal expense in this case so much greater than the wage? In moving from 10 to 11 workers, for example, the firm would have to pay one dollar more per hour to each of the 10 it originally planned to hire and then pay K9 to the added worker—for a total of K19 in extra costs. The marginal expense, then, includes the wages paid to the extra worker (as was the case in chapter 3) plus the additional cost of raising the wage for all other workers.6 The hypothetical data in Table 5.1 are graphed in Figure 5.2. The (solid) supply curve in Figure 5.2 indicates, of course, the number of employees attracted to the firm at each wage level. In short, it represents, for the firm in question, the wage it must pay to get to each of the employment levels it is considering. The dashed line represents the marginal expense—the added cost of increasing the employment level by one worker. The marginal expense curve both lies above the supply curve and is steeper in slope (that is, goes up at a faster rate).

Wage and Marginal Expense of Labour (K) Employment Supply Marginal Expense of Labour

**Figure 5.2**

**The Firm’s Choice of Wage and Employment Levels**

What are the labour market effects caused by having the marginal expense of labour lie above the wage rate?

To maximize profits, we know that any firm—including those in monopsonistic markets—should hire labour until the point at which the marginal revenue product of labour (MRPL) equals labour’s marginal expense (MEL): (5.1) To illustrate the effects of having MEL exceed the wage (W), we turn to Figure 5.3, which displays, for a given employer, its labour supply curve, the associated marginal expense of labour curve, and the downward-sloping curve depicting the firm’s MRPL. Any firm in a monopsonistic labour market must make two decisions about hiring. First, like firms in competitive labour markets, it must decide how much labour

to hire. This decision, consistent with the profit-maximizing criterion in equation (5.1), is made by finding the employment level at which. In Figure

5.3, the profit-maximizing level of employment for the firm shown is E\* because it is at E\* that (note the intersection of the relevant curves at point X).

Second, the firm must find the wage rate necessary to generate E\* employees. In Figure 5.3, the wage rate that will attract E\* workers is W\* (note point Y on the labour supply curve). The firm’s labour supply curve represents the relationship between its potential wage rates and the number of workers interested in working Profit-Maximizing Employment and Wage Levels in a Firm Facing a Monopsonistic Labour Market there. Thus, this second decision (about wages) is shown graphically by reading from the labour supply curve the wage needed to attract the profit-maximizing number of workers.

**Monopsonistic Conditions and Firms’ Wage Policies**

A difference between competitive and monopsonistic labour markets that immediately stands out concerns the wage policies of employers. With a competitive labour market, where individual firms are wage takers and can hire all the labour they want at the going wage, employers decide only on the number of workers they want to hire; the wage they pay is given to them by the market. We have seen, however, that firms facing monopsonistic conditions have a second decision to make: they must decide on the wage to pay as well. Further, while firms in competitive labour markets hire until the MRPL equals the (given) wage, firms in monopsonized markets pay workers a wage less than their marginal revenue product.

The implication that firms in monopsonistic labour r markets must have their own wage policies does not suggest, of course, that they set wages without constraints.

Thus, monopsonistic conditions do not give firms a completely free hand in

deciding on their wages; they must still face constraints imposed by both labour and product markets. Within the product and labour market constraints facing them, however, different firms in monopsonistic labour markets may well offer different wages to equivalent workers. It is unlikely that the labour supply and MRPL curves would be exactly the same for different firms in the same labour market; thus, we should not be surprised if exactly comparable workers were to have different marginal productivities and receive different wages at different firms. Thus, a firm employing older equipment and having a lower MRPL could coexist with one having new equipment and a higher MRPL by paying a lower wage to the same kind of

worker. Indeed, a careful summary of studies on wage differences and the law of one price found strong evidence suggesting that the same worker would receive different pay if he or she worked for different employers.

**How Do Monopsonistic Firms Respond to Shifts in the Supply Curve?**

In a monopsonistic labour market, the firm does not really have a labour demand curve! Labour demand curves for a firm are essentially derived from sequentially asking, “If the market wage were at some level (say, K5), what would be the firm’s profit-maximizing level of employment? If, instead, the wage was K6, what would be the firm’s desired level of employment?” Under monopsonistic conditions, the firm is not a wage taker, so asking hypothetical questions about the level of wages facing the firm is meaningless. Given the firm’s labour supply curve and its schedule of marginal revenue product (MRPL at various levels of employment), there is only one profit-maximizing level of employment and only one associated wage rate, both of which are chosen by the firm.

**Shifts in Labour Supply That Increase MEL**

Consider the short-run and long run effects on a monopsonistic firm’s desired level of employment if the supply curve facing the firm shifts (but remains upward-sloping). Suppose, for example, that the labour supply curve was to shift to the left, reflecting a situation in which fewer people are willing to work at any given wage level. With the competitive model of labour demand, a leftward shift of a market supply curve would cause the market wage to increase and the level of employment to fall, as employers moved to the left along their labour demand curves. Will these changes in wages and employment occur under monopsonistic conditions?

In Figure 5.4, the MRPL curve is fixed (we are in the short run), and the leftward shift of the labour supply curve is represented by a movement to curve from the original curve S. With a supply curve of S, the firm’s marginal expense of labour curve was MEL, and it chose to hire E workers and pay them a wage of W.

When the supply curve shifts to, the firm’s marginal labour expenses shift to a higher curve. Therefore, its new profit-maximizing level of employment

falls to E.g., and its new wage rate increases to W. Thus, with a monopsonistic

**Figure 5.4**

The Monopsonistic Firm’s Short-Run Response to a Leftward Shift in Labour Supply: Employment Falls and Wage Increases model (just as with the competitive model), a leftward shift in labour supply increases MEL, raises wages, and reduces firms’ desired levels of employment in the short run.

In the long run, labour’s increased marginal expense will induce the substitution of capital for labour as firms seek to find the cost-minimizing mix of capital and labour. You will recall that the cost-minimizing conditions for capital and labour under competitive conditions were given in equation (3.8c), in which the wage rate was treated as the marginal expense of labour. In a monopsonistic labour market, MEL exceeds W, so the left-hand side of equation (3.8c) must be written in its general form: (5.2)

Clearly, if a monopsonist is minimizing its costs of production and its MEL is increased, it will want to restore equality to condition (5.2) by substituting capital for labour. Thus, employment decreases even more in the long run than in the short run.

**Effects of a Mandated Wage**

Let us next consider what would happen if some non-market force were to compel the firm to pay a particular wage rate that was higher than the one it was paying. Would the firm’s desired level of employment decline? For a monopsonistic firm’s short-run response, refer to Figure 5.5, where the firm initially equates MRPL and MEL at point A and chooses to hire E0 workers,

which requires it to pay a wage of W0. Conditions: Both Wages and Employment Can Increase in the Short Run Suppose now that a mandated wage of Wm is set in Figure 5.5. This mandate prevents the firm from paying a wage less than Wm and effectively creates a horizontal portion (BD) in the labour supply curve facing the firm (which is now BDS).

The firm’s marginal expense of labour curve is now BDEM, because up to employment level E1, the marginal expense of labour is equal to Wm. The firm, which maximizes profits by equating marginal revenue with marginal expense (this equality is now at point C), will hire Em workers. Even though wages have risen from W0 to Wm, desired employment rises from E0 to Em!

For a monopsonistic firm, then, a mandated wage can simultaneously increase the average cost of labour (that is, the wages paid to workers) and reduce MEL. It is the decrease in marginal expense that induces the firm to expand output and employment in the short run. Thus, because an upward-sloping supply curve is converted to one that is horizontal, at least for employment near the current level, it is possible that both wages and employment can increase with the imposition of a mandated wage on a monopsonistic firm. This possibility is subject to two qualifications, however.

First, in the context of Figure 5.5, employment will increase only if the mandated wage is set between W0 and. A mandated wage above would increase MEL above its current level ( ) and cause the profit-maximizing level

of employment to fall below E0. (The student can verify this by drawing a horizontal line from any point above on the vertical axis and noting that it will intersect the MRPL curve to the left of E0.)

Second, Figure 5.5, with its fixed MRPL curve, depicts only the short-run

response to a mandated wage. In the long run, two (opposing) effects on employment are possible. With a mandated wage that is not too high, a monopsonistic firm’s MEL is reduced, causing a substitution of labour for capital in the long run. While the monopsonistic firm’s marginal expense of labour may have fallen, however, labour’s average cost (the wage) has increased. It is now more expensive to produce the same level of output than before; thus, profits will decline. If it is in a competitive product market, a firm’s initial profit level will be normal for that market, so the decline will push its profits below normal. Some owners will get out of the market, putting downward pressure on employment. If this latter (scale) effect is large enough, employment in monopsonistic sectors could fall in the long run if a mandated wage were imposed.

In summary, then, the presence of monopsonistic conditions in the labour

market introduces uncertainty into how employment will respond to the imposition of a mandated wage if the new wage reduces the firm’s marginal expense of labour. Any shift in the supply of labour curve that increases the marginal expense of labour, of course, will unambiguously reduce employment.

**Monopsonistic Conditions and the Employment Response to Minimum Wage Legislation**

Can the presence of monopsonistic conditions in the labour market offer a potential explanation for these findings?

We saw in the previous section that if the labour market is monopsonistic, legislated increases in the minimum wage raise wages but—if modest enough in size—can reduce the marginal expense of labour. Thus, our expectations about the direction of employment changes caused by a higher minimum wage are ambiguous: some firms might experience increases in employment (because MEL falls), but others might be forced to close because higher total labour costs render their operations nonprofitable. Our discussion in the previous section might also help explain why the labour demand elasticities presented in Table 4.1 tend to be larger (more elastic) than those implied by many studies of employment responses to minimum-wage changes.

The elasticities presented in Table 4.1 were estimated from wage and employment outcomes that were generated by market forces. Graphically, these estimates were derived from analyses like the one presented in Figure 5.4, where a leftward shift in the supply curve unambiguously caused wages to rise and employment to fall. Increases in the minimum wage cause a very different set of responses, as we saw when comparing Figures 5.4 and 5.5.

If monopsonistic conditions exist, then theory leads us to expect that employment responses to wage changes generated by market forces might be different from employment responses to legislated wage increases. Is it credible to assert that monopsonistic conditions might be what underlie the small or uncertain direction of employment changes we find in minimum wage studies? Most of these studies focus on teenagers, and one might think that teenagers could move almost without cost from one part-time job to another. If mobility is virtually costless for teenagers, they would freely move among employers in response to small wage differentials, the teenage labour market would correspond closely to the competitive model, and we would have to look elsewhere for an explanation of the uncertain estimated effects of minimum wages on teenage employment. We have argued that mobility is hindered (made costlier) by imperfect information about alternative wage offers and job requirements, by the time and aggravation of applying and being evaluated, and by the necessity of giving up valued nonwage job characteristics that might be difficult to replace in the new job. Teenagers, as well as adults, face these categories of cost. Moreover, teenagers often take jobs with the intent of staying only a short time, and they may perceive

the total gains from going to a higher-paying employer as too small to justify the investment of time and effort needed to change employers. Thus, it is not inconceivable that the supply curves to firms that typically employ teenagers (fast-food outlets, for example) are upward-sloping and that monopsonistic conditions prevail even in these places.

**Job Search Costs and Other Labour Market Outcomes**

The presence of job mobility costs for workers means that they must make decisions about when to search for a new employer (and incur the costs of search) and when to stay put. These decisions about search have some interesting implications that can help explain why wages rise with both labour market experience and the length of time (tenure) with a particular employer. Other reasons for why wages rise with experience and tenure will be discussed later in the text; however, our current discussion of job search costs warrants attention to these implications here. We will also discuss how job search costs affect decisions by those who are unemployed.

**Wage Levels, Luck, and Search**

We have seen that employee mobility costs can create monopsonistic conditions that result in pay differences among workers who have equal productive capabilities. Monopsonistic conditions, however, are not the sole cause of wage differences for workers who appear to be similar. Indeed, we will spend much time later in this text analysing wage differences associated with job or worker characteristics that are often not easily measured or observed: different working conditions, different on-the-job training requirements or opportunities, and different ways to use pay in creating incentives for productivity. In addition, we will also analyse wage differences related to racial, ethnic, or gender differences that may be unrelated to productive characteristics.

What the theory of monopsonistic labour markets offers to the analysis of wage differences, however, is the implication that to some extent, a worker’s

wage depends on luck. Some workers will be fortunate enough to obtain a job offer from a high-paying employer, and some will not. Furthermore, given the costs of changing employers, the mobility from low-wage to high-wage firms may never be great or rapid enough to bring wages into equality.

When workers who may think they can get improved job offers face costs in

searching for employers, we are naturally drawn to thinking about an employee–employer “matching” process that occurs over a period that may be lengthy. Workers can be viewed as wanting to obtain the best match possible but finding that there is a cost to getting better matches. Those who see their jobs as a poor match (perhaps because of low pay) have more incentives to search for other offers than do workers who are lucky enough to already have good matches (high wages). Over time, as the unlucky workers have more opportunity to acquire offers, matches for them should improve—but, of course, at some wage levels, likely wage increases from a search are so small (or, given the worker’s expected stay on the job, so short-lived) that further search is not worth the cost. Labour-market studies have observed that workers’ wages tend to increase both with (1) overall labour market experience and, (2) holding labour market experience constant, the length of time with one’s employer (“job tenure”).9 Job search considerations may play a role in producing these patterns, and we will briefly discuss them here.

**Wages and Labour Market Experience**

One of the things that make job search costly is that it takes time and effort to obtain job offers. Furthermore, job openings occur more or less randomly over time, so that during any one period in which a worker is “in the market,” not all potentially attractive openings even exist. As time passes, however, jobs open up and workers have a chance to decide whether to apply. Those who have spent more time in the labour market have had more chances to acquire better offers and thus improve upon their initial job matches. While other explanations are explored in chapter 9, the costs of job

search offer one explanation for why we observe that, in general, workers’ wages improve the longer they are active in the labour market.

**Wages and Job Tenure**

With costly job searches, workers who are fortunate enough to find jobs with high-paying employers will have little incentive to continue searching, while those who are less fortunate will want to search again.

This means that the workers who have been with their firms the longest will tend to be the ones who got higher wages to begin with, and we should therefore observe a positive correlation between tenure and earnings. Indeed, as noted above, empirical studies also find that among workers with the same skills and labour market experience, those who have longer job tenure with their employers also tend to have higher wages. While there are other potential explanations for this relationship as well (see chapters 9 and 11), the presence of costly job search suggests that it may not simply be longer tenures that cause higher wages; rather, higher wages can also cause longer job tenures!

**Job Search Costs and Unemployment**

Job search costs can also help to explain the existence (and level) of unemployment. While we analyse unemployment in chapter 14, the relationship between search costs and the phenomenon of unemployment is important to introduce at this point. Briefly put, searching for job offers is something that the unemployed must do, and the search process will take

time and effort. The longer it takes for a worker to receive an acceptable offer, the longer the unemployed worker will remain unemployed. Thus, higher job search costs will tend to lengthen the spells of unemployment and hence increase the unemployment rate.10

**Monopsonistic Conditions and the Relevance of the Competitive Model**

If employee mobility costs mean that monopsonistic conditions exist in the typical labour market, does this imply that the competitive model is irrelevant or misleading?

While we have seen that the competitive model does indeed offer predictions that are at least partially contradicted by the evidence, it is difficult to believe that it is irrelevant, especially in the long run. The major difference between the competitive and monopsonistic models, of course, is the assumption about employee mobility costs. When we consider workers as a group, however, mobility costs are likely to be higher in the near term than over the long haul. It is relatively costly, for example, for a registered nurse with a family established in Albany to move herself and her family to Sacramento.

Likewise, an established payroll clerk working with an employment agency may find it aggravating or time-consuming to search for, and then move to, a similar job in the furniture industry. It is much less costly, however, for a

recent graduate or immigrant who is trying to decide where in the country to

locate, or in which industry to work, to “move among” job offers. Recent graduates or immigrants have to search and make a decision anyway (established workers often do not), and when choosing among offers, they have much less to give up in terms of established relationships by taking one offer over the other. As time passes, those established in jobs retire and are replaced by new workers who see the advantages of locating in certain areas or accepting work in certain industries; thus, over time, we would expect wage differences owing to luck to dissipate— even if mobility costs are present in the short term. One study, for example, found that new immigrants to Zambia are more likely to be clustered in states offering the highest wages for their skill groups and that their presence has

helped to narrow regional wage differences. It is also the case that, monopsonistic conditions notwithstanding, employers cannot deviate too far from the market when setting wages, for if they do, they will encounter problems in attracting, retaining, and motivating their workers. Nobel laureate Paul Samuelson put the issue this way in his bestselling economics textbook: Just because competition is not 100 per cent perfect does not mean that it must be zero. The world is a blend of (1) competition and (2) some degree of monopoly power over the wage to be paid. A firm that tries to set its wage too low will soon learn this. At first, nothing much need happen; but eventually, it will find its workers quitting a little more rapidly than would otherwise be the case. Recruitment of new people of the same quality will get harder and harder, and slackening off in the performance and productivity of those who remain on the job will become noticeable.

**Frictions on the Employer Side of the Market**

Employers also face frictions in searching for and hiring employees. These frictions cause firms to bear costs that are associated with the number of workers hired rather than the hours they work, and they are called “quasi-fixed” costs because they are either difficult or impossible to cut in the short run—unlike variable costs (such as hourly wages), which can be readily cut by reducing the hours of work. The presence of quasi-fixed costs slows the adjustment of employment levels to changing market conditions faced by firms. The types of quasi-fixed costs are first discussed in this section, and we then move to an analysis of their implications for the labour market behaviour of firms.

**Categories of Quasi-Fixed Costs**

Employers often incur substantial quasi-fixed costs in hiring and compensating their employees. In general, these costs fall into two categories: investments in their workforce and certain employee benefits. We discuss each type of quasi-fixed costs below.

**In summary**

**Activity**

**1.** Define the monopsonistic labour markets

2. Describe profit maximisation under monopsonistic conditions

3. Discuss wages and labour market experience

4. Explain why the marginal expenses of labour exceed the wage rate.

5. Describe job search costs and unemployment

6. Discuss the firm’s choice of wage and employment levels.

7. Describe how monopsonistic firms respond to shift in the supply curve

8. Discuss the monopsonistic conditions and the employment response to minimum wage legislation

9. Explain job search costs and other Labour market outcomes

**CHAPTER TEN**

**LABOUR INVESTMENTS**

**OBJECTIVES**

At the end of the chapter, students should be able to:

1. Discuss labour investment
2. Describe hours devoted by firms to training a new worker a new worker during first three months on job
3. Explain employee benefits
4. Explain the training decision by employers
5. Describe the types of training
6. Explain employer training investments and recessionary lay offs
7. Discuss the use of credentials and internal labour markets
8. Describe how the employer can recoup its hiring investments

**INTRODUCTION**

When an employer has a job vacancy, it must incur certain costs in finding a suitable employee to hire. It has to advertise the position, screen applications, interview potential candidates, and (in the case of highly sought applicants) “wine and dine” the worker selected. A 1982 survey, for example, which was weighted toward employers hiring less-skilled workers, found that even for these vacancies, almost 22 person-hours were spent screening and interviewing applicants. Once hired, there are the additional costs of orienting the new worker and getting him or her on the payroll.

A hiring cost not to be overlooked—especially because it has been the subject of public policy debates—is the cost of terminating the worker. Every

employee a firm hire might also have to be let go if economic circumstances or job performance require it. As we discuss in Example 5.1, policies that require severance pay or otherwise increase the costs of ending the employment relationship thus add to the quasi-fixed costs of hiring workers.

**Table 5.2**

**Hours Devoted by Firms to Training a New Worker during First Three**

**Months on Job**

**Activity Average Hours**

Hours of formal instruction by training personnel 19 Hours spent by management in orientation, informal training, extra supervision 59 Hours spent by co-workers in informal training 34 Hours spent by new worker watching others do work 41 Total 153

Source: John Bishop, “The Incidence of and Payoff to Employer Training,” Cornell University Centre for Advanced Human Resource Studies Working Paper 94–17, July 1994, 11.

In addition to the hiring costs, firms typically provide formal or informal training to both their new and continuing workers. The costs of this training generally fall into three classes:

1. The explicit monetary costs of formally employing trainers and providing

training materials.

2. The implicit, or opportunity, costs of lost production incurred when experienced employees take time to demonstrate procedures to trainees’ in

less-formal settings.

1. The implicit, or opportunity, costs of the trainee’s time. A survey in the early 1990s found that in the first three months (or 520 hours of work) an employee is with a firm, about 30 percent (153 hours) of his or her time is spent in training. The data from this study, summarized in **Table 5.2**, also suggest that very little of this training was formal classroom-type instruction; most took place informally at the workstation.

Hiring and training costs can be categorized as investments because they are incurred in the present and have benefits (in the form of increased productivity) only in the future. Investments are inherently risky because, once made, the costs are “sunk,” and there are no guarantees about future returns.

**Employee Benefits**

Besides their direct wage and salary earnings, workers also typically receive nonwage compensation in the form of employer-provided medical and life insurance, retirement plans, vacation days, Social Security payments, and other employee benefits. Table 5.3 details the employee benefits received by workers in 2010, and it is important to note that many of these benefits represent quasi-fixed costs to the employer. That is, many employee benefits are associated with the number of employees but not with the hours they work.

Most life and medical insurance policies have premiums to the employer that are charged on a per-worker basis and are not proportional to the hours worked. Pay for time not worked (vacation, holidays, and sick leave) also tends to be quasi-fixed. Some pension costs are proportional to hours worked because many employers offer defined contribution plans and make payments to a retirement fund for each worker that are proportional to wage or salary earnings. However, some employers have defined benefit pension plans that promise pension payments to retirees that are a function of years of service, not hours of work; the costs of these plans are thus quasi-fixed in nature.

In the category of legally required benefits, workers’ compensation insurance

costs are strictly proportional to hours worked, because they are levied as a

percentage of payroll, and Social Security taxes are proportional for most employees. 15 However, the unemployment insurance payroll-tax liability is specified to be a percentage (the tax rate) of each employee’s yearly earnings up to a maximum level (the taxable wage base), which in 2010 was between K7,000 and K15,000 in over two-thirds of all states.16 Since most employees earn more than K15,000 per year, having an employee work an additional hour per week will not cause any increase in the employer’s payroll-tax liability. Therefore, unemployment insurance costs are a quasi-fixed cost to most employers.

**In Table 5.3,** we have indicated (by a superscript a) which nonwage costs are usually of a quasi-fixed nature. The data suggest that around 19 percent of total compensation (about 60 percent of nonwage costs) is quasi-fixed. These quasi fixed costs averaged, on a yearly basis, over K10,600 per worker in 2010. The quasi-fixed nature of many nonwage labour costs has important effects on employer hiring and overtime decisions. These effects are discussed in the following section.

**The Employment/Hours Trade-Off**

The simple model of the demand for labour presented in the preceding chapters spoke to the quantity of labour demanded, making no distinction between the number of individuals employed by a firm and the average length of its employees’ workweek. Holding all other inputs constant, however, a firm can produce a given level of output with various combinations of the number of employees hired and the number of hours worked per week. Presumably, increases in the number of employees hired will allow for shorter workweeks, whereas longer workweeks will allow for fewer employees, other things equal.

In chapter 3, we defined the marginal product of labour (MPL) as the change

in output generated by an added unit of labour, holding capital constant. Once we distinguish between the number of workers hired (which we will denote by M) and the hours each works on average (H), we must think of two marginal products of labour. MPM is the added output associated with an added worker, holding both capital and average hours per worker constant. MPH is the added output generated by increasing average hours per worker, holding capital and the number of employees constant. As with MPL, we assume that both MPM and MPH are positive but that they decline as M and H (respectively) increase. How does a firm determine its optimal employment/hours combination? Is it ever rational for a firm to work its existing employee’s overtime on a regularly scheduled basis, even though it must pay them a wage premium, rather than hiring additional employees?

**Determining the Mix of Workers and Hours**

The fact that certain labour costs are not hours-related, while others are, will lead employers to think of “workers” and “hours-per-worker” as two substitutable labour inputs. Thus, the profit-maximizing employer will weigh the cost of producing an added unit of output by hiring more workers against the cost of producing an added unit of output by employing its current workers for more hours. Recalling our discussion of equation 3.8c,

profit maximization can only be achieved when these two costs are equal. Thus, if the marginal expense of hiring an added worker is MEM, and the marginal expense of hiring current workers for an extra hour is MEH, then for profits to be maximized, the following condition must hold:

(5.3)

The left-hand side of equation (5.3) is the cost of an added unit of output produced by hiring more workers, and the right-hand side is the cost of an added unit of output produced by hiring workers for more hours.

One implication of equation (5.3) is that if for some reason MEM rises relative to MEH, firms will want to substitute hours for workers by hiring fewer employees but having each work more hours. (An alternative to hiring more workers or increasing hours is to “rent” workers; see Example 5.2.) Conversely, if MEH rises relative to MEM, the employer will want to produce its profit-maximizing level of output with a higher ratio of workers to average hours per worker.

The relationship between and hours of work is graphed in Figure 5.6, which indicates that as MEM rises relative to MEH, other things equal, hours of

work per employee tend to rise.

**Policy Analysis: The Overtime Pay Premium**

In Zambia, the Fair Labour Standards Act requires that employees covered by the act (generally, hourly paid, nonsupervisory workers) receive an overtime pay premium of at least 50 percent of their regular hourly wage for each hour worked in excess of 40 per week. Many overtime hours are worked because of unusual circumstances that MEM>MEH=MEHMPH Weekly Overtime (hours per employee) are difficult or impossible to meet by hiring more workers: rush orders, absent workers, and mechanical failures are all examples of these emergency situations.

However, some overtime is regularly scheduled; for example, over 20 percent of men who are skilled craft workers or technicians usually work more than 44 hours per week.18 Given the “time-and-one-half” premium that must be paid for overtime work, we can conclude that employers who regularly schedule overtime do so because it is cheaper than incurring the quasi-fixed costs of employing more workers. Indeed, the production workers most likely to work long hours on a regular basis are those for which hiring and training costs are higher. For example, while over 20 percent of male craft workers are scheduled for more than 44 hours each week, only 12 percent of unskilled males usually work more than 44 hours. 19

In the fall of 2004, the Zambian Ministry of Labour introduced several controversial revisions to the government overtime regulations that redefined which jobs are exempt from coverage. Generally speaking, for a job to be exempt from the requirements of overtime pay, the employee must be paid on a salaried basis (not by the hour) and perform administrative, professional, or executive duties. The regulations introduced in 2004 disallowed exemptions for low-paying salaried jobs (paying less than K455 per week), regardless of duties—thus adding overtime coverage to an estimated 1.3 million workers. The new regulations, however, revised the definitions of “administrative,” “professional,” and “executive” duties and added many computer and outside sales jobs to the list of those exempt from overtime regulations. Also made exempt were jobs in which total pay exceeds K100, 000 per year.20

These revisions created a storm of public comment and criticism. While they

were lauded for giving “greater protection” to low-paid hourly employment, the revisions were also criticized for making it easier to exempt jobs, thus “making it likely that millions of [workers] will work longer hours at reduced pay.

**Overtime and Spreading the Work**

It is often argued that the time-and-one-half requirement for overtime protects workers by “spreading the work” (creating more job openings) through reduced usage of overtime. One reason to be cautious in our expectations that increased coverage will create more jobs is that applying the overtime premium increases the average cost of labour even if a firm eliminates its prior use of overtime! Firms using overtime before could have increased their workforce and reduced the use of overtime earlier; the fact that they did not suggests that the quasi-fixed costs of hiring made that a costlier option. If they now eliminate overtime and hire more workers at the same base wage rate, their labour costs will clearly rise. Increased labour costs will tend to reduce both the scale of output and increase firms’ incentives to substitute capital for labour, thereby reducing the total labour hours demanded by affected firms. Thus, even if base wages are not changed, it is unlikely that all the reduced overtime hours will be replaced by hiring more workers.

**Overtime and Total Pay**

Will newly covered workers experience an increase in earnings, and will those in newly exempt jobs experience some earnings decrease as a result of the revisions? It is possible that they will not, because the base wage rate may change in response to changes in overtime coverage. We have seen that many overtime hours are regularly scheduled, and in these cases, it is possible that employers and employees mutually agree (informally, at least) on a “package” of weekly hours and total compensation. If so, firms that regularly schedule overtime hours might respond to a legislated increase in

coverage by reducing the straight-time salary in a way that, after taking the newly required overtime payments into account, would leave total compensation per worker unchanged. Similarly, employees who lose coverage under overtime laws and are asked to work more hours may be unwilling to stay in those jobs—unless, of course, their pay is increased accordingly.

Thus, the long-run effects of overtime regulations on the total earnings of

workers may not be as profound as supporters imply. Put differently, in firms that paid above-average overtime premiums, straight-time (base) wages were below average—and firms that paid above-average base wages paid below-average overtime premiums.

**Training Investments**

We have identified employer-provided training as an important investment that can increase the quasi-fixed costs of hiring workers. The costs of training, even if provided by the employer, are often at least partly paid by workers themselves in one way or another, so training investments represent a rather unique friction in the labour market. This section explores the implications of this friction for both employer behaviour and employee behaviour.

**The Training Decision by Employers**

Consider an employer who has just hired a new employee. If the employer

decides to bear the cost of training this worker, it will incur the explicit and

implicit training costs discussed earlier—including, of course, the forgone output of the worker being trained. Thus, in the training period, the employer is likely to be bearing costs on behalf of this new worker that are greater than the worker’s marginal revenue product. Under what conditions would an employer be willing to undertake this kind of investment?

As with any investment, an employer that bears net costs during the training period would only do so if it believes that it can collect returns on that investment after training. It is the prospect of increased employee productivity that motivates an employer to offer training, but the only way the employer can make a return on its investment is to “keep” some of that added post-training revenue by not giving all of it to the worker in the form of a wage increase. Put succinctly, for a firm to invest in training, two conditions must be met.

First, the training that employees receive must increase their marginal revenue productivity more than it increases their wage. Second, the employees must stay with the employer long enough for the employer to receive the required returns (obviously, the longer the employees stay with the firm, other things equal, the more profitable the investment will be).

**The Types of Training**

At the extremes, there are two types of training that employers can provide.

General training teaches workers skills that can be used to enhance their productivity with many employers; learning how to speak English, use a word-processing program, drive a truck, or create Web sites are examples of general training. At the other end of the spectrum is specific training, which teaches workers skills that increase their productivity only with the employer providing the training. Examples of specific training include teaching workers how to use a machine unique to their workplace or orienting them to particular procedures and people they will need to deal with in various circumstances they will encounter at work.

**General Training**

Paying for general training can be a rather risky investment for an employer, for if the employer tries to keep post-training wage increases below increases in marginal revenue productivity, trained workers might leave. Because general training raises productivity with other employers too, trained workers have incentives to look for higher wage offers from employers that have no training costs to recoup! Thus, if employee mobility costs are not very great, employers will be deterred from investing in general training. The likelihood of making back their required returns is low, because the gap between marginal revenue product and the post-training wage might not be sufficiently great, or the expected tenure of the trained workers with the firms sufficiently long, to recoup their investment costs. When worker-mobility costs are low, firms either would not provide the training or would require the employees to pay for it by offering a very low (or, in the case of some interns, a zero) wage rate during the training period.

Only if employees are deterred from quitting by high mobility costs does our theory suggest that firms would invest in general training. Recent work suggests that firms often do invest in general training for their workers, and these investments are cited as yet another reason for believing that the labour market is characterized by monopsonistic conditions.24 Put succinctly, for a firm to invest in training, two conditions must be met.

First, the training that employees receive must increase their marginal revenue productivity more than it increases their wage. Second, the employees must stay with the employer long enough for the employer to receive the required returns (obviously, the longer the employees stay with the firm, other things equal, the more profitable the investment will be). characterized by monopsonistic conditions.

**Specific Training**

Employers have stronger incentives to invest in specific training, because such training does not raise the worker’s productivity with other firms, and it therefore does not make the worker more attractive to competing employers. While the training itself does not increase the outside offers an

employee might be able to receive, a firm undertaking investments in specific training must nevertheless take precautions to keep the trained employee from quitting, because once the employee quits, the employer’s investment is destroyed (that is, returns on the investment cannot be realized). Thus, concerns about the possibility that trained employees will quit before the employer can receive its required investment returns exist relative to specific, as well as general, training. These concerns lead us to a discussion of (a) who bears the costs of training and (b) the size of post-training wage increases.

**Training and Post-Training Wage Increases**

Consider a situation in which worker-mobility costs are relatively low, and the employer is considering bearing all the costs of training. With investment costs to recoup, the employer would be unable to raise wages very much after training and still have incentives to invest. We know that higher wages reduce the probability of a worker quitting, so by failing to increase the wage much after training, the employer would put its investment at risk. Trained workers might decide to quit at even a small provocation (the boss is in a bad mood one day, for example, or they are asked to work overtime for a while), and without some assurance that trained employees will stay, the firm would be reluctant to make a training investment for which it bore all the costs.

Conversely, if a firm’s employees paid for their own training by taking a

lower wage than they could get elsewhere during the training period, they would require the benefits of a much higher post-training wage to make employment at the firm attractive. If they were to get their entire improved marginal revenue product in the form of a wage increase, however, an employer that finds it relatively inexpensive to hire and fire workers would have little to lose by firing them at the smallest provocation—and if they get fired, their investment is destroyed!

Thus, if labour market frictions are otherwise small, the best way to provide

incentives for on-the-job training is for employers and employees to share the costs and returns of the investment. If employees pay part of these costs, the post training wage can be increased more than if employers bear all the training costs—and the increased post-training wage protects firms’ investments by reducing the chances trained workers will quit. The training costs borne by employers must be recouped by not raising the post-training wage very much, but this condition helps protect workers’ investments by making it attractive for firms to retain them unless the provocation is major (we discuss the issue of layoffs in more detail a bit later in this chapter). Put differently, if both employers and employees share in the costs of training, and thus share in the returns, they both have something to lose if the employment relationship is ended in the post-training period.

Empirical studies measuring the wage profiles associated with on-the-job

training in the United States, however, suggest that employers bear much of the costs and reap most of the returns associated with training. Wages apparently are not depressed enough during the training period to offset the employer’s direct costs of training, so subsequent wages increases are much smaller than productivity increases. A survey of employers, summarized in Figure 5.7, estimated that productivity increases, which generally rose with the hours of initial on-the-job training, were far larger than wage increases over a worker’s first two years with an employer. Other studies that directly link the wage profiles of American workers with the amount of training they have received find that post-training wage increases are relatively modest. 26 The evidence that employers bear much of the training costs, and reap much of the returns, suggests that these employers believe their workers face relatively high worker-mobility costs. These firms are willing to bear the investment costs because they do not feel the need to raise the post-training wage much in order to retain their trained employees.

**Employer Training Investments and Recessionary Layoffs**

We have seen that employers will have incentives to invest in worker training only when the post-training marginal revenue productivity is expected to be sufficiently above the wage so that the investment returns are attractive. Suppose a firm has made the investment but at some point thereafter finds that its workers’ marginal revenue productivity falls below what it expected because of a business downturn (a “recession”). If it cannot lower wages for one reason or another, will the firm want to lay off its trained workers?

In general, firms will not want to lay off their workers as long as the workers

are bringing in revenues that are in excess of their wages. Even if the gap

between marginal revenue productivity and wage is not sufficient to yield an

attractive return on the firm’s training investment, those training costs—once incurred—are “sunk.” While the firm might wish it had not invested in training, the best it can do after training is get what returns it can. Workers who are laid off clearly bring in no returns to the employer, so its incentives are to retain any worker whose marginal revenue productivity exceeds his or her wage. Of course, if the downturn causes marginal revenue productivity to still fall below the wage rate, firms do have incentives to lay off trained workers (unless they believe the downturn will be very short and do not want to take the risk that the laid-off workers will search for other employment).

The presence of employer training investments, then, offers an explanation for two phenomena we observe in the labour market. First, as a general rule, we observe that workers who are least susceptible to being laid off during recessions are the most skilled and those with the longest job tenures. Older and more skilled workers are those most likely to have been the objects of past employer training investments, and they therefore tend to enter recessions with larger gaps between marginal revenue product and wage. These gaps cushion any fall in marginal revenue product and provide their employers with stronger incentives to keep on employing them during the downturn. Workers who enter the recession with wages closer to marginal revenue productivity are more likely to find that the downturn causes their marginal revenue product to fall below their wage, and when this occurs, employers may find it profitable to lay them off.

Second, we observe that average labour productivity—output per labour hour—falls in the early stages of a recession and rises during the early stages of recovery. As demand and output start to fall, firms that have invested in worker training respond by keeping their trained workers on the payroll even though their marginal productivity falls. Such “labour hoarding” causes output per worker to fall. Of course, when demand picks up again, firms can increase output without proportionately increasing their employment because, in effect, they have maintained an inventory of trained labour. In the latter situation, output per worker rises.

**Hiring Investments**

In addition to training employees, firms must also evaluate them when making hiring, placement, and promotion decisions. They may therefore find that training programs—even ones with a “general” component—can be used to help them discover the learning abilities, work habits, and motivation levels of new employees (see Example 5.3).28 Thus, some of what appears to be general training may actually represent an investment in firm-specific information about employees that will be useful later on in making assignments and deciding on promotions.

We conclude this chapter with a section that analyses hiring and screening investments in greater detail.

**The Use of Credentials**

Since firms often bear the costs of hiring and training workers, it is in their interest to make these costs as low as possible. Other things equal, firms should prefer to obtain a workforce of a given quality at the least possible cost. Similarly, they should prefer to hire workers who are fast learners, because such workers could graduate may be fully qualified to work for a firm that insists on college graduates.

Excluding them from the pool of potential applicants imposes costs on them

(they do not get the job); however, it also imposes costs on the employer if other qualified applicants cannot be readily found. On the other hand, there may be some unproductive workers among the group of college graduates, and an employer who hires them may well suffer losses while they are employed. However, if the reduction in hiring costs that arises when signals (such as educational credentials, marital status, or age) are used is large, it may prove profitable for an employer to use them even if an occasional unsatisfactory worker sneaks through.

**Internal Labour Markets**

One of the difficulties in hiring employees is that such personal attributes as

dependability, motivation, honesty, and flexibility are difficult to judge from interviews, employment tests, or even the recommendations of former employers. This difficulty has led many larger firms to create an internal labour market, in which workers are hired into relatively low-level jobs and higher-level jobs are filled only from within the firm. This policy gives employers a chance to observe actual productive characteristics of the employees hired, and this information is then used to determine who stays with the firm and how fast and how high employees are promoted. The benefits of using an internal labour market to fill vacancies are that the firm knows a lot about the people working for it. Hiring decisions for upper-level

jobs in either the blue-collar or the white-collar workforces will thus offer few surprises to the firm. The costs of using the internal labour market are associated with the restriction of competition for the upper-level jobs to those in the firm. Those in the firm may not be the best employees available, but they are the only ones the firm considers for these jobs. Firms most likely to decide that the benefits of using an internal labour market outweigh the costs are those whose upper-level workers must have a lot of firm-specific knowledge and training that can best be attained by on-the-job learning over the years.

As noted earlier, firms that pay for training will want to ensure that they

obtain employees who can learn quickly and will remain with them long enough for the training costs to be recouped through the post-training surplus. For these firms, the internal labour market offers two attractions. First, it allows the firm to observe workers on the job and thus make better decisions about which workers will be the recipients of later, perhaps very expensive, and training. Second, the internal labour market tends to foster an attachment to the firm by its employees. They know that they have an inside track on upper-level vacancies because outsiders will not be considered. If they quit the firm, they would lose this privileged position. They are thus motivated to become long-term employees of the firm.

The full implications of internal labour markets for wage policies within the firm will be discussed in chapter 11.

**How Can the Employer Recoup Its Hiring Investments?**

Whether a firm invests in training its workers or in selecting them, it will do so only if it believes it can generate an acceptable rate of return on its investment. For a labour investment to be worthwhile, an employer must be able to benefit from a situation in which workers are paid less than their marginal value to the firm in the post-investment period. How can employers generate a post-investment surplus from their hiring investments?

Suppose that applicants for a job vacancy have average, below-average, or

above-average productivity but that the employer cannot tell which without making some kind of investment in acquiring that information. If the firm does not make this investment, it must assume that any particular applicant is of average ability and pay accordingly. If the firm makes an investment in acquiring information about its applicants, however, it could then hire only those whose productivity is above average. The surplus required to pay back its investment costs would then be created by paying these above-average workers a wage less than their true productivity.

Would the firm pay its new workers the average wage even though they are

above average in productivity, thereby obtaining the full surplus? As with the case of training, the firm would probably decide to pay a wage greater than the average, but still below workers’ actual productivity, to increase the likelihood that the workers in whom it has invested will remain. If its workers quit, the firm would have to invest in acquiring information about their replacements. While the self-interest of employers would drive them to pay an above-average wage to above-average workers, two things could allow the screening firm to pay a wage that is still lower than workers’ full productivity. One is the presence of mobility costs among employees. The other is that information one employer finds out through a costly screening process may not be observable by other employers without an investment of their own. Either of these conditions would inhibit employees from obtaining wage offers from competing firms that could afford to pay full-productivity wages because they had no screening expenses to recoup. How do worker-mobility costs affect the slope of labour supply curves to individual firms?

2. Why do upward-sloping labour supply curves to firms cause the marginal

expense of labour to exceed the wage rate?

3. One recent magazine article on economic recovery from a recession argued: “Labour productivity growth usually accelerates in the first year of an expansion, because firms are slow to hire new labour.” Comment.

4. “Minimum wage laws help low-wage workers because they simultaneously

increase wages and reduce the marginal expense of labour.” Analyse this statement.

5. An author recently asserted: “Low-wage jobs provide fewer hours of work than high-wage jobs.” According to economic theory, is this statement likely to be correct? Why?

6. Workers in a certain job are trained by the company, and the company calculates that to recoup its investment costs, the workers’ wages must be K5 per hour below their marginal productivity. Suppose that after training, wages are set at K5 below marginal productivity but that developments in the product market quickly (and permanently) reduce marginal productivity by K2 per hour. If the company does not believe it can lower wages or employee benefits, how will its employment level be affected in the short run? How will its employment level be affected in the long run? Explain, being sure to define what you mean by the short run and the long run!

7. For decades, most large employers bought group health insurance from insurers who charged them premiums on a per-worker basis. In 1993, a proposal for a national health insurance plan contained a provision requiring group health insurers to charge premiums based on payroll (in effect, financing health insurance by a payroll “tax”). Assuming the total premiums paid by employers remain the same, what are the labour market implications of this proposed change in the way in which health insurance is financed?

8. The manager of a major league baseball team argues: “Even if I thought Player X was washed up, I couldn’t get rid of him. He’s in the third year of a four-year, K24-million deal. Our team is in no position financially to eat the rest of his contract.” Analyse the manager’s reasoning by using economic theory.

9. The president of France has announced that his government is considering abandoning

**In summary**

**Activity**

1. Discuss labour investment

2. Describe hours devoted by firms to training a new worker a new worker during first three months on job

3. Explain employee benefits

4. Explain the training decision by employers

5. Describe the types of training

6. Explain employer training investments and recessionary lay offs

7. Discuss the use of credentials and internal labour markets

8. Describe how the employer can recoup its hiring investments

**CHAPTER ELEVEN**

**SUPPLY OF LABOUR TO THE ECONOMY: THE DECISION TO WORK**

OBJECTIVES

At the end of the chapter, students should be able to:

1. Discuss the supply of labour to the economy; the decision to work
2. Explain the trends in labour force participation and hours of work
3. Explain labour force participation rates
4. Describe a theory of the decision to work
5. Explain the opportunity cost of leisure
6. Discuss wealth and income
7. Discuss the income and substitution effects.

**INTRODUCTION**

This chapter will focus on issues of worker behaviour. Labour supply decisions can be roughly divided into two categories that includes decisions about whether to work at all and, if so, how long to work. Questions that must be answered include whether to participate in the labour force, whether to seek part-time or full-time work, and how long to work both at home and for pay. The second category of decisions deal with the questions that must be faced by a person who has decided to seek work for pay: the occupation or general class of occupations in which to seek offers. This chapter begins with some basic facts concerning labour force participation rates and hours of work. We then develop a theoretical framework that can be used in the analysis of decisions to work for pay. This framework is also useful for analysing the structure of various income maintenance programs.

**Trends in Labour Force Participation and Hours of Work**

When a person actively seeks work, he or she is, by definition, in the labour force. As pointed out in chapter 2, the labour force participation rate is the percentage of a given population that either has a job or is looking for one. Thus, one clear-cut statistic important in measuring people’s willingness to work outside the home is the labour force participation rate.

**Labour Force Participation Rates**

One of the most dramatic changes in the labour market over the past six decades has been the increased labour force participation of women, especially married women. Table 6.1 shows the dimensions of this change. As recently as 1950, less than 25 percent of married women were in the labour force, but by 1980, this percentage had doubled. Recently, the labour force participation rate of married women has reached over 60 percent, although since 2000, the growth for married women seems to have stopped and the rates for single women have fallen.1 One interest of this chapter is in understanding the forces underlying these changes.

**Table 6.1**

**Hours of Work**

Because data on labour force participation include both the employed and those who want a job but do not have one, they are a relatively pure measure of labour supply. In contrast, the weekly or yearly hours of work put in by the typical employee are often thought to be determined only by the demand side of the market. After all, don’t employers, in responding to the factors discussed in chapter 5, set the hours of work expected of their employees? They do, of course, but hours worked are also influenced by employee preferences on the supply side of the market, especially in the long run.

Even though employers set work schedules, employees can exercise their

preferences regarding hours of work through their choice of part-time or fulltime work, their decisions to work at more than one job, or their selection of occupations and employers.2 For example, women managers who work full-time average more hours of work per week than full-time clerical workers, and male sales workers work more hours per week than their full-time counterparts in skilled craft jobs. Moreover, different employers offer different mixes of full-time and part-time work, require different weekly work schedules, and have different policies regarding vacations and paid holidays.

Employer offers regarding both hours and pay are intended to enhance their profits, but they must also satisfy the preferences of current and prospective

employees. For example, if employees receiving an hourly wage of $X for 40 hours per week really wanted to work only 30 hours at $X per hour, some enterprising employer (presumably one with relatively lower quasi-fixed costs) would eventually seize on their dissatisfaction and offer jobs with a 30-hour workweek, ending up with a more satisfied, productive workforce in

the process. While the labour supply preferences of employees must be satisfied in the long run, most of the short-run changes in hours of work seem to emanate from the demand side of the market.3 Workweeks typically vary over the course of a business cycle, for example, with longer hours worked in periods of robust demand. In analysing trends in hours of work, then, we must carefully distinguish between the forces of supply and demand.

In the first part of the twentieth century, workers in Zambia manufacturing plants typically worked 55 hours per week in years with strong economic activity; in the last two decades, American manufacturing workers have worked, on average, less than 40 hours per week during similar periods. For example, in the years 1988, 1995, and 2004—when the unemployment rate was roughly 5.5 percent and falling—manufacturing production workers averaged 38.4, 39.3, and 38.6 hours per week, respectively. In general, the decline in weekly hours of manufacturing work in Zambia occurred prior to 1980, and since then, hours of work have shown little tendency to decline.4

**A Theory of the Decision to Work**

Can labour supply theory help us to understand the long-run trends in labour force participation and hours of work noted above? Because labour is the most abundant factor of production, it is fair to say that any country’s well-being in the long run depends heavily on the willingness of its people to work. Leisure and other ways of spending time that do not involve work for pay are also important in generating well-being; however, any economy relies heavily on goods and services produced for market transactions. Therefore, it is important to understand the work-incentive effects of higher wages and incomes, different kinds of taxes, and various forms of income maintenance programs.

The decision to work is ultimately a decision about how to spend time. One

way to use our available time is to spend it in pleasurable leisure activities. The other major way in which people use time is to work. We can work around the home, performing such household production as raising children, sewing, building, or even growing food. Alternatively, we can work for pay and use our earnings to purchase food, shelter, clothing, and child care. Because working for pay and engaging in household production are two

ways of getting the same jobs done, we shall initially ignore the distinction

between them and treat work activities as working for pay. We shall therefore be characterizing the decision to work as a choice between leisure and working for pay. Most of the crucial factors affecting work incentives can be understood in this context, but insight into labour supply behaviour can also be enriched by a consideration of household production; this we do in chapter 7.

If we regard the time spent eating, sleeping, and otherwise maintaining ourselves as more or less fixed by natural laws, then the discretionary time we have (16 hours a day, say) can be allocated to either work or leisure. It is most convenient for us to begin our analysis of the work/leisure choice by analysing the demand for leisure hours.

**Some Basic Concepts**

Basically, the demand for a good is a function of three factors:

1. The opportunity cost of the good (which is often equal to market price). 2. One’s level of wealth.

3. One’s set of preferences. For example, consumption of heating oil will vary with the cost of such oil; as that cost rises, consumption tends to fall unless one of the other two factors intervenes. As wealth rises, people generally want larger and warmer houses that obviously require more oil to heat.5 Even if the price of energy and the level of personal wealth were to remain constant, the demand for energy could rise if a falling birth rate and lengthened life span resulted in a higher proportion of the population being aged and therefore wanting warmer houses. This change in the composition of the population amounts to a shift in the overall preferences for warmer houses and thus leads to a change in the demand for heating oil. (Economists usually assume that preferences are given and not subject to immediate change. For policy purposes, changes in prices and wealth are of paramount importance in explaining changes in demand because these variables are more susceptible to change by government or market forces.)

**Opportunity Cost of Leisure**

To apply this general analysis of demand to the demand for leisure, we must first ask, “What is the opportunity cost of leisure?” The cost of spending an hour watching television is basically what one could earn if one had spent that hour working. Thus, the opportunity cost of an hour of leisure is equal to one’s wage rate—the extra earnings a worker can take home from an extra hour of work.6

**Wealth and Income**

Next, we must understand and be able to measure wealth. Naturally, wealth includes a family’s holdings of bank accounts, financial investments, and physical property. Workers’ skills can also be considered assets, since these skills can be, in effect, rented out to employers for a price. The more one can

get in wages, the larger the value of one’s human assets. Unfortunately, it is not usually possible to directly measure people’s wealth. It is much easier to measure the returns from that wealth, because data on total income are readily available from government surveys. Economists often use total income as an indicator of total wealth, since the two are conceptually so closely related.7

**Defining the Income Effect**

Theory suggests that if income increases while wages and preferences are held constant, the number of leisure hours demanded will rise. Put differently, if income increases, holding wages constant, desired hours of work will go down. (Conversely, if income is reduced while the wage rate is held constant, desired hours of work will go up.) Economists call the response of desired hours of leisure to changes in income, with wages held constant, the income effect. The income effect is based on the simple notion that as incomes rise, holding leisure’s opportunity cost constant, and people will want to consume more leisure (which means working less).

Because we have assumed that time is spent either in leisure or in working

for pay, the income effect can be expressed in terms of the supply of working hours as well as the demand for leisure hours. Because the ultimate focus of this chapter is labour supply, we choose to express this effect in the context of supply. Using algebraic notation, we define the income effect as the change in hours of work ( ) produced by a change in income ( ), holding wages constant(6.1)

We say the income effect is negative because the sign of the fraction in equation (6.1) is negative. If income goes up (wages held constant), hours of work fall. If income goes down, hours of work increase. The numerator ( ) and denominator ( ) in equation (6.1) move in opposite directions, giving a negative sign to the income effect.

**Defining the Substitution Effect**

Theory also suggests that if income is held constant, an increase in the wage rate will raise the price and reduce the demand for leisure, thereby increasing work incentives. (Likewise, a decrease in the wage rate will reduce leisure’s opportunity cost and the incentives to work, holding income constant.) This substitution effect occurs because as the cost of leisure changes, income held constant, leisure and work hours are substituted for each other. In contrast to the income effect, the substitution effect is positive. Because this

effect is the change in hours of work ( ) induced by a change in the wage ( ),

holding income constant , the substitution effect can be written as (6.2) Because the numerator ( ) and denominator ( ) always move in the same direction, at least in theory, the substitution effect has a positive sign.

**Observing Income and Substitution Effects Separately**

At times, it is possible to observe situations or programs that create only one effect or the other. (Laboratory experiments can also create separate income and substitution effects; an experiment with pigeons, discussed in Example 6.1, suggests that labour supply theory can even be generalized beyond humans!) Usually, however, both effects are simultaneously present, often working against each other. Receiving an inheritance offers an example of the income effect by itself. The bequest enhances wealth (income) independent of the hours of work. Thus, income is increased without a change in the compensation received from an hour of work. In this case, the income effect induces the person to consume more leisure, thereby reducing the willingness to work. (Some support for this theoretical prediction can be seen later in Example 6.3.) Observing the substitution effect by itself is rare. Social Security revenues are collected by a tax on earnings, so reductions in the tax are, in effect, increases in the wage rate for most workers. For the average person, however, the increased wealth associated with this wage increase would have been exactly offset by increases in the gasoline tax. Hence, wages would have been increased while income was held more or less constant. This program would have created a substitution effect that induced people to work more hours.

**Both Effects Occur When Wages Rise**

While the above examples illustrate situations in which the income or the substitution effect is present by itself, normally both effects are present, often working in opposite directions. The presence of both effects working in opposite directions creates ambiguity in predicting the overall labour supply response in many cases. Consider the case of a person who receives a wage increase.

The labour supply response to a simple wage increase will involve both an income effect and a substitution effect. The income effect is the result of the worker’s enhanced wealth (or potential income) after the increase. For a given level of work effort, he or she now has a greater command over resources than before (because more income is received for any given number of hours of work).

The substitution effect results from the fact that the wage increase raises the opportunity costs of leisure. Because the actual labour supply response is the sum of the income and substitution effects, we cannot predict the response in advance; theory simply does not tell us which effect is stronger.

If the income effect is stronger, the person will respond to a wage increase by decreasing his or her labour supply. This decrease will be smaller than if the same change in wealth were due to an increase in non-labour wealth, because the substitution effect is present and acts as a moderating influence. However, as seen in Example 6.2, when the income effect dominates, the substitution effect is not large enough to prevent labour supply from declining. It is entirely plausible, of course that the substitution effect will dominate. If so, the actual response to wage increases will be to increase labour supply. Should the substitution effect dominate, the person’s labour supply curve— relating, say, his or her desired hours of work to wages—will be positively sloped. That is, labour supplied will increase with the wage rate. If, on the other hand, the income effect dominates, the person’s labour supply curve will be negatively sloped.

Economic theory cannot say which effect will dominate, and in fact, individual labour supply curves could be positively sloped in some ranges of the wage and negatively sloped in others. In Figure 6.1, for example, the person’s desired hours of work increase (substitution effect dominates) when wages go up as long as wages are low (below W\*). At higher wages, however, further increases result in reduced hours of work (the income effect dominates); economists refer to such a curve as backward-bending.

**Analysis of the Labour/Leisure Choice**

This section introduces indifference curves and budget constraints—visual aids that make the theory of labour supply easier to understand and to apply to complex policy issues. These graphical aids visually depict the basic factors underlying the demand for leisure (supply of labour) discussed earlier.

**Preferences**

Let us assume that there are two major categories of goods that make people happy—leisure time and the goods people can buy with money. If we take the prices of goods as fixed, then they can be compressed into one index that is measured by money income (with prices fixed, more money income means Wage it is possible to consume more goods). Using two categories, leisure and money income, allows our graphs to be drawn in two-dimensional space. Since both leisure and money can be used to generate satisfaction (or utility), these two goods are to some extent substitutes for each other. If forced to give up some money income—by cutting back on hours of work, for example—some increase in leisure time could be substituted for this lost income to keep a person as happy as before.

To understand how preferences can be graphed, suppose a thoughtful consumer/worker were asked to decide how happy he or she would be with a daily income of K64 combined with 8 hours of leisure (point a in Figure 6.2). This level of happiness could be called utility level A. Our consumer/worker could name other combinations of money income and leisure hours that would also yield utility level A. Assume that our respondent named five other combinations. All six combinations of money income and leisure hours that yield utility level A are represented by heavy dots in Figure 6.2. The curve connecting these dots is called an indifference curve, which connects the various combinations of money income and leisure that yield equal utility. (The term indifference curve is derived from the fact that since each point on the curve yields equal utility, a person is truly indifferent about where on the curve he or she will be.)

Our worker/consumer could no doubt achieve a higher level of happiness if

he or she could combine the 8 hours of leisure with an income of K100 per day instead of just K64 a day. This higher satisfaction level could be called utility level B. The consumer could name other combinations of money income and leisure that would also yield this higher level of utility. These combinations are denoted by the Xs in Figure 6.2 that are connected by a second indifference curve. Indifference curves have certain specific characteristics that are reflected in the way they are drawn:

Utility level B represents more happiness than level A. Every level of leisure consumption is combined with a higher income on B than on A. Hence, our respondent prefers all points on indifference curve B to any point on curve A. A whole set of indifference curves could be drawn for this one person, each representing a different utility level. Any such curve that lies to

the northeast of another one is preferred to any curve to the southwest because the north eastern curve represents a higher level of utility.

Indifference curves do not intersect. If they did, the point of intersection would represent one combination of money income and leisure that yielded two different levels of satisfaction. We assume our worker/ consumer is not so inconsistent in stating his or her preferences that this could happen.

Indifference curves are negatively sloped because if either income or leisure hours are increased, the other is reduced in order to preserve the same level of utility. If the slope is steep, as at segment LK in Figure 6.3, a given loss of income need not be accompanied by a large increase in leisure hours to keep utility constant.9 When the curve is relatively flat, however, as at segment MN in Figure 6.3, a given decrease in income must be accompanied by a large increase in the consumption of leisure to hold utility constant. Thus, when indifference curves are relatively steep, people do not value money income as highly as when such curves are relatively flat; when they are flat, a loss of income can only be compensated for by a large increase in leisure if utility is to be kept constant.

Indifference curves are convex—steeper at the left than at the right. This shape reflects the assumption that when money income is relatively high and leisure hours are relatively few, leisure is more highly valued (and income less valued) than when leisure is abundant and income relatively scarce. At segment LK in Figure 6.3, a great loss of income (from Y4 to Y3, for example) can be compensated for by just a little increase in leisure, whereas a little loss of leisure time (from H3 to H4, for example) would require a relatively large increase in income to maintain equal utility.

What is relatively scarce is more highly valued.

**Figure 6.3**

An Indifference Curve Conversely, when income is low and leisure is abundant (segment MN in Figure 6.3), income is more highly valued. Losing income (by moving from Y2 to Y1, for example) would require a huge increase in leisure for utility to remain constant. To repeat, what is relatively scarce is assumed to be more highly valued.

Finally, different people have different sets of indifference curves. The curves drawn in Figures 6.2 and 6.3 were for one person. Another person would have a completely different set of curves. People who value leisure more highly, for example, would have had indifference curves that were generally steeper (see Figure 6.4a). People who do not value leisure highly would have relatively flat curves (see Figure 6.4b). Thus, individual preferences can be portrayed graphically.

**Income and Wage Constraints**

Everyone would like to maximize his or her utility, which would be ideally done by consuming every available hour of leisure combined with the highest conceivable income. Unfortunately, the resources anyone can command are limited. Thus, all that is possible is to do the best one can, given limited resources. To see these resource limitations graphically requires superimposing constraints on one’s set of indifference curves to see which combinations of income and leisure are available and which are not.

Suppose the person whose indifference curves are graphed in Figure 6.2 had no source of income other than labour earnings. Suppose, further, that he or she could earn K8 per hour.

Figure 6.5 includes the two indifference curves shown in Figure 6.2 as well as a straight line (ED) connecting combinations of leisure and income that are possible for a person with an K8 wage and no outside income. If 16 hours per day are available for work Money Income per Day

**(a) Person Who Places High Value on an Extra Hour of Leisure**

**(b) Person Who Places Low Value on an Extra Hour of Leisure**

Now / is exactly the slope of the budget constraint (in absolute value).11

Figure 6.5 shows how the constraint raises K8 for every 1-hour increase in work: if the person works 0 hours, income per day is zero; if the person works 1 hour, K8 in income is received; if he or she works 5 hours, K40 in income is achieved. The constraint raises K8 because the wage rate is K8 per hour. If the person could earn K16 per hour, the constraint would rise twice as fast and therefore be twice as steep. It is clear from Figure 6.5 that our consumer/worker cannot achieve utility level B. He or she can achieve some points on the indifference curve representing utility level A—specifically, those points between L and M in Figure 6.5. However, if our consumer/worker is a utility maximizer, he or she will realize that a utility level above A is possible. Remembering that an infinite number of indifference curves can be drawn between curves A and B in Figure 6.5, one representing each possible level of satisfaction between A and B, we can draw a curve ( ) that is northeast of curve A and just tangent to the budget constraint at point N. Any movement along the budget constraint away from the tangency point places the person on an indifference curve lying below A¿. Workers who face the same budget constraint, but who have different preferences for leisure, will make different choices about hours of work. If the person whose preferences were depicted in Figure 6.5 had placed lower values on leisure time—and therefore had indifference curves that were comparatively flatter, such as the one shown in Figure 6.4b—then the point of tangency with constraint ED would have been to the left of point N (indicating more hours of work). Conversely, if he or she had steeper indifference curves, signifying that leisure time was more valuable (see Figure 6.4a), then the point of tangency in Figure 6.5 would have been to the right of point N, and fewer hours of work would have been desired. Indeed, some people will have indifference curves so steep (that is, preferences for leisure so strong) that there is no point of tangency with ED. For these people, as is illustrated by Figure 6.6, utility is maximized at the “corner” (point D); they desire no work at all and therefore are not in the labour force.

**The Income Effect**

Suppose now that the person depicted in Figure 6.5 receives a source of income independent of work. Suppose further that this non-labour income amounts to about K36 per day. Thus, even if this person worked 0 hours per day, his or her daily income would be K36. Naturally, if the person worked more than 0 hours, his or her daily income would be equal to K36 plus earnings (the wage multiplied by the hours of work).

Our person’s command over resources has clearly increased, as can be shown by drawing a new budget constraint to reflect the non-labour income. As shown by the darker blue line in Figure 6.7, the endpoints of the new constraint are point d (0 hours of work and K36 of money income) and point e (16 hours of work and K164 of income—K36 in non-labour income plus K128 in earnings). Note that the new constraint is parallel to the old one. Parallel lines have the same slope; since the slope of each constraint reflects the wage rate, we can infer that the increase in non-labour income has not changed the person’s wage rate.

We have just described a situation in which a pure income effect should be observed. Income (wealth) has been increased, but the wage rate has remained unchanged. The previous section noted that if wealth increased and the opportunity cost of leisure remained constant, the person would consume more leisure and work less.

**Income and Substitution Effects with a Wage Increase**

Suppose that instead of increasing one’s command over resources by receiving a source of non-labour income, the wage rate was to be increased from K8 to K12 per hour. This increase, as noted earlier, would cause both an income effect and a substitution effect; workers would be wealthier and face a higher opportunity cost of leisure. Theory tells us in this case that the substitution effect pushes them toward more hours of work and the income effect toward fewer, but it cannot tell us which effect will dominate. Figures 6.8 and 6.9 illustrate the possible effects of the above wage change on a person’s labour supply, which we now assume is initially 8 hours per day.

Figure 6.8 illustrates the case in which the observed response by a worker is to increase the hours of work; in this case, the substitution effect is stronger than the income effect. Figure 6.9 illustrates the case in which the income effect is stronger and the response to a wage increase is to reduce the hours of work. The difference between the two figures lies solely in the shape of the indifference curves that might describe a person’s preferences; the budget constraints, which reflect wealth and the wage rate, are exactly the same.

Figures 6.8 and 6.9 both show the old constraint, AB, the slope of which

reflects the wage of K8 per hour. They also show the new one, AC, which reflects Money Income (dollars)

**Isolating Income and Substitution Effects**

Is it possible to graphically isolate the substitution effect? The answer is yes, and the most meaningful way to do this is to return to the context of a wage change, such as the one depicted in Figures 6.8 and 6.9. Overall effect of a wage increase on the labour supply of the person whose preferences are depicted. As we saw earlier, the effect of the wage increase in this case is to raise the person’s utility from U1 to U2 and to induce this worker to increase desired hours of work from 8 to 11 per day. Embedded in this overall effect of the wage increase, however, is an income effect pushing toward less work and a substitution effect pushing toward more.

By definition, the income effect is the change in desired hours of work brought on by increased wealth, holding the wage rate constant. To reveal this embedded effect, we ask a hypothetical question: “What would have been the change in labour supply if the person depicted in panel

1. had reached the new indifference curve (U2) with a change in non-labour income instead of a change in his or her wage rate?”

We begin to answer this question graphically by moving the old constraint

to the northeast, which depicts the greater command over leisure time and

goods—and hence the higher level of utility—associated with greater wealth. The constraint is shifted outward while maintaining its original slope (reflecting the old $8 wage), which holds the wage constant. The dashed line in panel

(b), which is parallel to AB, depicts this hypothetical movement of the old constraint, and it results in a tangency point at N3. This tangency suggests that had the person received non-labour income, with no change in the wage, sufficient to reach the new level of utility, he or she would have reduced work hours from 8 (N1) to 7 (N3) per day. This shift is graphical verification that the income effect is negative, assuming that leisure is a normal good.

The substitution effect is the effect on labour supply of a change in the wage

rate, holding wealth constant. It can be seen in panel (c) of Figure 6.10 as the difference between where the person actually ended up on indifference curve U2 (tangency at N2) and where he or she would have ended up with a pure income effect (tangency at N3). Comparing tangency points on the same indifference curve is a graphical approximation to holding wealth constant. Thus, with the wage change, the person represented in Figure 6.10 ended up at point N2, working 11 hours a day. Without the wage change, the person would have chosen to work 7 hours a day (point N3). The wage change by itself, holding utility (or real wealth) constant, caused work hours to increase by 4 per day. This increase demonstrates that the substitution effect is positive.

To summarize, the observed effect of rising wages from K8 to K12 per hour

increased the hours of work in Figure 6.10 from 8 to 11 per day. This observed effect, however, is the sum of two component effects. The income effect, which operates because a higher wage increases one’s real wealth, tended to reduce the hours of work from 8 to 7 per day. The substitution effect, which captures the pure effect of the change in leisure’s opportunity cost, tended to push the person toward 4 more hours of work per day. The end result was an increase of 3 in the hours worked each day.

**Which Effect Is Stronger?** Suppose that a wage increase changes the budget constraint facing a worker from CD to CE in Figure 6.11. If the worker had a relatively flat set of indifference curves, the initial tangency along CD might be at point A, implying a relatively heavy work schedule. If the person had more steeply sloped indifference curves, the initial tangency might be at point B, where hours at work are fewer.

One important influence on the size of the income effect is the extent of the

northeast movement of the new constraint: the more the constraint shifts outward, the greater the income effect will tend to be. For a person with an initial tangency at point A, for example, the northeast movement is larger than that for a person whose initial tangency is at point B. Put in words, the increased command over resources made possible by a wage increase is only attainable if one works, and the more work-oriented the person is, the greater will be his or her increase in resources. Other things equal, people who are working longer hours will exhibit greater income effects when wage rates change.

To take this reasoning to the extreme, suppose a person’s indifference curves were so steep that the person was initially out of the labour force (that is, when the budget constraint was CD in Figure 6.11, his or her utility was maximized at point C). The wage increase and the resultant new constraint, CE, can induce only two outcomes:

the person will either begin to work for pay or remain out of the labour force. Reducing the hours of paid employment is not possible. For those who are

out of the labour force, then, the decision to participate as wage offers rise clearly reflects a dominant substitution effect. Conversely, if someone currently working decides to change his or her participation decision and drop out of the labour force when wages fall, the substitution effect has again dominated. Thus, the labour force participation decisions brought about by wage changes exhibit a dominant substitution effect. We turn now to a more detailed analysis of the decision whether to join the labour force.

**The Reservation Wage**

An implication of our labour supply theory is that if people who are not in the labour force place a value of KX on the marginal hour of leisure, then they would be unwilling to take a job unless the offered wages were greater than KX. Because they will “reserve” their labour unless the wage is KX or more (see Example 6.4), economists say that they have a reservation wage of KX. The reservation wage, then, is the wage below which a person will not work, and in the labour/leisure context, it represents the value placed on an hour of lost leisure time.14

**Empirical Findings on the Income and Substitution Effects**

Labour supply theory suggests that the choices workers make concerning their desired hours of work depend on their wealth and the wage rate they can command, in addition to their preferences. In particular, this theory suggests the existence of a negative income effect and a positive substitution effect. Empirical tests of labour supply theory generally attempt to determine if these two effects can be observed, if they operate in the expected directions, and what their relative Most recent studies of labour supply have used large samples of individuals to analyse how labour force participation and hours of work are affected by wage rates and income, holding other influences (age, for example) constant. Studies of male and female labour force behaviour are done separately because of the different

roles men and women typically play in performing household work and childrearing— activities that clearly affect labour supply decisions but about which information is usually very limited.

The studies of labour supply behaviour for men between the ages of 25 and 55 generally conclude that both income and substitution effects are small, perhaps even zero. Probably because the net responses to wage changes are so close to zero, the results of studies that try to separately measure the income and substitution effects—while generally supportive of the theory—are highly dependent on the statistical methods used.

Studies of the labour supply behaviour of married women generally have

found a greater responsiveness to wage changes than is found among men, and recent work suggests two generalizations. First, changes in the hours of work associated with a wage change for married women are closer to those for men than are changes in labour force participation; that is, as seen in Example 6.5, the labour force participation rate for married women has been more responsive to wage changes than have been the hours of work. Second, in the last two decades, the labour supply behaviour of married women has become much more similar to that for men—meaning that the labour supply of women is becoming less responsive to wage changes than it used to be. The reduced responsiveness has been especially noticeable in women’s labour force participation decisions, where the differences between men and women have been greatest.

This growing similarity in labour supply behaviour may well reflect a growing similarity in the expectations held by women and men concerning work and careers.

**Programs with Net Wage Rates of Zero**

The programs just discussed were intended to confer benefits on those who are unable to work, and the budget-constraint spike was created by the eligibility requirement that to receive benefits, one must not be working. Other social programs, such as welfare, have different eligibility criteria and calculate benefits differently. These programs factor income needs into their eligibility criteria and then pay benefits based on the difference between one’s actual earnings and one’s needs. We will see that paying people the difference between their earnings and their needs creates a net wage rate of zero; thus, the work-incentive problems associated with these welfare programs result from the fact that they increase the income of program recipients while also drastically reducing the price of leisure.

**Nature of Welfare Subsidies**

Welfare programs have historically taken the form of a guaranteed annual income, under which the welfare agency determines the income needed by an eligible person (Yn in Figure 6.14) based on family size, area living costs, and local welfare regulations. Actual earnings are then subtracted from this needed level, and a check is issued to the person each month for the difference.

If the person does not work, he or she receives a subsidy of Yn. If the person

works, and if earnings cause dollar-for-dollar reductions in welfare benefits,

then a budget constraint like ABCD in Figure 6.14 is created. The person’s income remains Yn as long as he or she is subsidized. If receiving the subsidy, then, an extra hour of work yields no net increase in income, because the extra earnings result in an equal reduction in welfare benefits.

shift creates an income effect tending to reduce labour supply from the hours associated with point E to those associated with point F. However, it also causes the wage to effectively drop to zero; every dollar earned is matched by a dollar reduction in welfare benefits. This dollar-for-dollar reduction in benefits induces a huge substitution effect, causing those accepting welfare to reduce their hours of work to zero (point B). Of course, if a person’s indifference curves were sufficiently flat so that the curve tangent to segment CD passed above point B (see Figure 6.15), then that person’s utility would be maximized by choosing work instead of welfare.

**Welfare Reform**

In light of the disincentives for work built into traditional welfare programs, the United States adopted major changes to its come-subsidy programs in the 1990s. The Personal Responsibility and Work Opportunity Reconciliation Act (PRWORA) of 1996 gave states more authority over how they could design their own welfare programs, with the intent of leading to more experimentation in program characteristics aimed at encouraging work, reducing poverty, and moving people off welfare. PRWORA also placed a five-year (lifetime) time limit on the receipt of welfare benefits and required that after two years on welfare, recipients must work at least 30 hours per week. These changes appear to have had the effect of increasing the labour force participation rates of single mothers (the primary beneficiaries of the old welfare system); the participation rate for single mothers jumped from 68 percent in 1994 to roughly 78 percent in 2000—a much larger increase than was observed for other groups of women.

**Life time Limits**

Both lifetime limits and work requirements can be analysed using the graphical tools developed in this chapter. Lifetime limits on the receipt of welfare have the effect of ending eligibility for transfer payments, either by forcing recipients off welfare or by inducing them to leave so they can “save” their eligibility in case they need welfare later in life. Thus, in terms of Figure 6.14, the lifetime limit ultimately removes ABC from the potential recipient’s budget constraint, which then reverts to the market constraint of AD. Clearly, the lifetime limit increases work incentives by ultimately eliminating the income subsidy. However, within the limits of their eligible years, potential welfare recipients must choose when to receive the subsidy and when to “save” their eligibility in the event of a future need. Federal law provides for welfare subsidies only to families with children under the age of 18; consequently, the closer one’s youngest child is to 18 (when welfare eligibility ends anyway), the smaller are the incentives of the parent to forgo the welfare subsidy and save eligibility for the future.25

**Work Requirements**

As noted earlier, PRWORA introduced a work requirement into the welfare system, although in some cases, unpaid work or enrolling in education or training programs counts toward that requirement. States differ in how the earnings affect welfare benefits, and many have rules that allow welfare recipients to keep most of what they earn (by not reducing, at least by much, their welfare benefits); we analyse such programs in the next section. For now, we can understand the basic effects of a work requirement by maintaining our assumption that earnings reduce welfare benefits dollar for dollar.

Figure 6.16 illustrates the budget constraint associated with a minimum work requirement of 6 hours a day (30 hours per week). If the person fails to work the required 6 hours a day, no welfare benefits are received, and he or she will be along segment AB of the constraint. If the work requirement is met, but earnings are less than Yn, welfare benefits are received (see segment BCD). If the work requirement is exceeded, income (earnings plus benefits) remains at Yn—the person is along CD—until earnings rise above needed income and the person is along segment DE of the constraint and no longer eligible for welfare benefits.

The work-incentive effects of this work requirement can be seen from analysing Figure 6.16 in the context of people whose skills are such that they are potential welfare recipients. At one extreme, some potential recipients may have such steeply sloped indifferences curves (reflecting a strong preference, or a need, to stay at home) that utility is maximized along segment AB, where so little market work is performed that they do not qualify for welfare. At the opposite extreme, others may have such flat indifference curves (reflecting a strong preference for income and a weak preference for leisure) that their utility is maximized along segment DE; they work so many hours that their earnings disqualify them for welfare benefits.

In the middle of the above extremes will be those whose preferences lead

them to work enough to qualify for welfare benefits. Clearly, if their earnings

reduce their benefits dollar for dollar—as shown by the horizontal segment DC in Figure 6.16—they will want to work just the minimum hours needed to qualify for welfare, because their utility will be maximized at point C and not along DC.

(For **Subsidy Programs with Positive Net Wage Rates**

So far, we have analysed the work-incentive effects of income maintenance programs that create net wage rates for program recipients that are either negative or zero (that is, they create constraints that have either a spike or a horizontal segment). Do these programs offer a solution to the problem of work incentives? We will answer this question by analysing a relatively recent and rapidly growing program: the Earned Income Tax Credit (EITC).

The EITC program makes income tax credits available to low-income families

with at least one worker. A tax credit of K1 reduces a person’s income taxes by K1, and in the case of the EITC, if the tax credit for which workers qualify exceeds their total income tax liability, the government will mail them a check for the difference. Thus, the EITC functions as an earnings subsidy, and because the subsidy goes only to those who work, the EITC is seen by many as an income maintenance program that preserves work incentives. This view led Congress to vastly expand the EITC under President Bill Clinton and it is now the largest cash subsidy program directed at low-income households with children. The tax credits offered by the EITC program vary with one’s earnings and the number of dependent children. For purposes of our analysis, which is intended to illustrate the work-incentive effects of the EITC, we will focus on the credits in the year 2009 offered to unmarried workers with two children. Figure 6.17 graphs the relevant program characteristics for a worker with two children who could earn a market (unsubsidized) wage reflected by the slope of AC. As we will see later, for such a worker, the EITC created a budget constraint of ABDEC.

For workers with earnings of K12, 570 or less, the tax credit was calculated at 40 percent of earnings. That is, for every dollar earned, a tax credit of 40 cents was also earned; thus, for those with earnings of under K12, 570, net wages (Wn) were 40 percent higher than market wages (W). Note that this tax credit is represented by segment AB on the EITC constraint in Figure 6.17 and that the slope of AB exceeds the slope of the market constraint AC.

The maximum tax credit allowed for a single parent with two children was

K5, 028 in 2009. Workers who earned between K12, 570 and K16, 420 per year qualified for this maximum tax credit. Because these workers experienced no increases or reductions in tax credits per added dollar of earnings, their net wage is equal to their market wage. The constraint facing workers with earnings in this range is represented by segment BD in Figure 6.17, which has a slope equal to that of segment AC

**In summary**

**Activity**

1. Discuss the supply of labour to the economy; the decision to work

2. Explain the trends in labour force participation and hours of work

3. Explain labour force participation rates

4. Describe a theory of the decision to work

5. Explain the opportunity cost of leisure

6. Discuss wealth and income

7. Discuss the income and substitution effects.